

GESIS

ZUMA

NACHRICHTEN

58

Mai 2006

Zentrum für Umfragen, Methoden und Analysen (ZUMA)

ZUMA ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS). Die GESIS ist eine Einrichtung der *Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz* (WGL) (www.wgl.de).

Vorsitzender des Trägervereins ZUMA e.V.: Prof. Dr. Jan van Deth

Direktor: Prof. Dr. Peter Ph. Mohler

Hausanschrift

B 2, 1
68 159 Mannheim

Postanschrift

Postfach 12 21 55
68 072 Mannheim

Telefon

0621/1246 - 0

Fax

0621/1246 - 100

Redaktion: 0621/1246 - 228

E-Mail

porst@zuma-mannheim.de

Internet

www.gesis.org/zuma ZUMA
www.gesis.org/ GESIS

Herausgeber: PD Dr. Christof Wolf

Redaktion: Rolf Porst

ISSN 0941-1670 30. Jahrgang

Die ZUMA-Nachrichten veröffentlichen – neben Nachrichten des Instituts – Artikel, die von Interesse für die empirische Sozialforschung, insbesondere die Umfrageforschung, sind. Alle Artikel müssen einen methodischen Fokus haben oder zumindest methodische Aspekte stark betonen. Das Spektrum möglicher Artikel ist breit: Es reicht von Grundlagenforschung über angewandte Papiere bis zu Arbeiten, die einen praktisch-operativen Charakter haben. Die Artikel in den ZUMA-Nachrichten sollen für eine breite Leserschaft von Wissenschaftlern und Praktikern im Bereich der empirischen Sozialforschung verständlich sein. Alle Beiträge, die zur Veröffentlichung in den ZUMA-Nachrichten eingereicht werden, werden von mindestens zwei unabhängigen Gutachtern blind begutachtet.

Namentlich gekennzeichnete Beiträge geben die Meinung der Autoren wieder. Der Nachdruck von Beiträgen ist nach Absprache möglich. Die ZUMA-Nachrichten erscheinen im Mai und November eines Jahres. Sie werden Interessenten auf Anforderung kostenlos zugesandt. Die ZUMA-Nachrichten finden Sie auch im www unter der Adresse: http://www.gesis.org/publikationen/zeitschriften/zuma_nachrichten/.

Druck: PrintArt GmbH Druckerei + Verlag, Dannstadt/Mannheim
Gedruckt auf chlorfrei gebleichtem Papier.

© **ZUMA, Mannheim, Mai 2006**

Editorial.....	5
Nachruf auf Karl Schuessler <i>Dagmar Krebs & Manfred Kuchler</i>	7
Nachruf auf Philip James Stone <i>Ekkehard Mochmann</i>	10

FORSCHUNGSBERICHTE

Der Rolling Cross-Section Survey – ein Instrument zur Analyse dynamischer Prozesse der Einstellungsentwicklung. Bericht zur ersten deutschen RCS-Studie anlässlich der Bundestagswahl 2005 <i>Rüdiger Schmitt-Beck, Thorsten Faas & Christian Holst</i>	13
Overreporting: Ein vernachlässigtes Problem und die Schwierigkeit von Gegenmaßnahmen <i>Sibylle Hardmeier & Marie-Christine Fontana</i>	50
Zur Wirkung unterschiedlicher materieller Incentives in postalischen Befragungen. Ein Literaturbericht <i>Fred Berger</i>	81

MITTEILUNGEN

Computerunterstützte Vercodung der International Standard Classification of Occupations (ISCO-88): Vorstellen eines Instruments <i>Jürgen H.P. Hoffmeyer-Zlotnik, Doris Hess & Alfons J. Geis</i>	101
1. Nutzerkonferenz EVS „Forschung mit der Einkommens- und Verbrauchsstichprobe (EVS): Analysen zur sozialen Ungleichheit“.....	114
Gesammelte Erfahrung für alle verfügbar: Handbuch der Berufsvercodung im Netz.....	116
Call for Proposals: Themenschwerpunkt „Politische Partizipation, politische Kultur, Sozialkapital“ im ALLBUS 2008.....	118

Bericht über das zweite Treffen der Arbeitsgruppe MOBILSAMPLE. Tagungsankündigung	120
Zur Zukunft der Statistikausbildung. Forderungen aus der Praxis an die Ausgestaltung der Bachelor- und Master-Studiengänge der Wirtschafts- und Sozialwissenschaften. Pressemitteilung DESTATIS vom 9. Februar 2006.....	121
Zur Zukunft der Statistikausbildung Pressemitteilung DESTATIS vom 5. April 2006..	125

BUCHBESPRECHUNGEN

Gordon B. Willis: Cognitive Interviewing. A Tool for Improving Questionnaire Design, 2004.....	128
Nina Bauer: Verlaufsmusteranalyse – Methodologische Konsequenzen der Zeitlichkeit sozialen Handelns, 2005.....	131

ZUMA-PUBLIKATIONEN

Neue Bücher bei ZUMA.....	133
ZUMA-Nachrichten Spezial.....	135
ZUMA-Arbeitsberichte	137
ZUMA-Methodenberichte	139

ZUMA-VERANSTALTUNGEN 2006

ZUMA-Workshops: Zweites Halbjahr 2006.....	141
--	-----

NACHRICHTEN AUS DEM INSTITUT	142
---	-----

DURCHWAHL-RUFNUMMERN	144
-----------------------------------	-----

EDITORIAL

Liebe Leserin, lieber Leser,

seit einigen Jahren können Sie jedes Heft der ZUMA-Nachrichten sofort nach seinem Erscheinen auf unserer Homepage auch digitalisiert im pdf-Format einsehen und herunterladen. Dieser Service stand bisher nur für die Hefte ab Ausgabe 38 (Mai 1996) zur Verfügung. Seit kurzem liegen nun alle Hefte der ZUMA-Nachrichten einheitlich in digitalisierter Form vor. Alle Dokumente sind textlich voll erschlossen und können nach Begriffen oder Namen durchsucht werden. Zur besseren Recherche einzelner Beiträge steht Ihnen außerdem ein Gesamtinhaltsverzeichnis aller bisher erschienen Hefte zur Verfügung. Die elektronischen Ausgaben der Zeitschrift finden Sie unter: http://www.gesis.org/Publikationen/Zeitschriften/ZUMA_Nachrichten/index.htm

Mannheim, im Mai 2006

CHRISTOF WOLF

NACHRUF AUF KARL SCHUESSLER

15. Februar 1915 bis 26. Dezember 2005



Karl Schuessler, „Distinguished Professor“ an der Indiana University in Bloomington, Indiana, war einer der Großen in der amerikanischen Soziologie. Seine akademische Ausbildung wurde – nach einem Master von der University of Chicago im Jahre 1939 – durch den zweiten Weltkrieg unterbrochen, und er verbrachte vier Jahre bei der Marine. Nach Kriegsende setzte er seine Arbeiten an der Dissertation als Schüler von Edwin H. Sutherland fort und promovierte 1947 an der IU in Bloomington

mit einer Arbeit über die Wirkung des sozioökonomischen Hintergrunds auf musikalischen Geschmack.

Nach einem Jahr als Instructor an der Vanderbilt University in Nashville, kehrte Karl Schuessler an die Indiana University zurück und begann dort seine außergewöhnliche Karriere vom Assistant Professor bis schließlich zum Distinguished Professor (1976). Karl Schuessler hat die Indiana University nie verlassen, hat jedoch viel Zeit als Gastprofessor im Ausland (China, Thailand, Indien, Deutschland) verbracht und war in der „sociological community“ außerordentlich engagiert auf nationaler ebenso wie auf internationaler Ebene.

Schon in seiner Dissertation werden die beiden Säulen in Karl Schuesslers Leben deutlich – seine Liebe zur Musik und sein Engagement für die Soziologie. In der Soziologie richtete sich Karls Interesse wesentlich auf die Methoden der Empirischen Sozialforschung, wobei sein Hauptanliegen stets darin bestand, eine praxisrelevante Sozialforschung zu betreiben, deren Ergebnisse eine nützliche Basis für politische Entscheidungen sein könnten. Seine zentrale Frage bei allen Aktivitäten war: „Does this make the world a better place to live?“. Methodische Spitzfindigkeiten ohne praktischen Bezug waren für ihn irrelevant.

In seiner Zeit als Dekan der Faculty for Sociology an der Indiana University in den sechziger Jahren hatte Karl Schuessler maßgeblichen Anteil an der Gründung des dortigen „Institute for Social Research“ und der Etablierung einer praxisbezogenen Ausbildung der

Studenten. Wie kein anderer verstand er es, methodologische Innovation und Akkuratessse mit inhaltlich gehaltvoller Forschung zu verknüpfen. Und dabei war er nicht einseitig auf die quantitative Forschung festgelegt, sondern sein Anliegen bestand ganz im Weber'schen Sinne darin, zu „verstehen“. Seine methodologische Rigorosität, gepaart mit seiner Offenheit für eine Vielzahl von Problemsichten und Zugangsweisen, machten ihn zu einem idealen Kandidaten für das prestigeträchtige Amt des Herausgebers des *American Sociological Review* – ein Amt, das er von 1969-1971 inne hatte – und des jährlich erscheinenden Sammelbandes *Sociological Methodology* (von 1978-1980).

Er hat zahlreiche Bücher und Artikel zu methodologischen und statistischen Fragestellungen veröffentlicht. Das auch in Deutschland weit verbreitete Buch über „Social Data Analysis“ und vielleicht mehr noch das grundlegende Lehrbuch „Statistical Reasoning in Sociology“ (zuerst 1961 mit John Mueller, dann in späteren Auflagen mit Herbert Costner) bringen deutlich zum Ausdruck, wie Statistik anschaulich und praxisrelevant dargestellt werden kann und wie wichtig Denken in quantitativen Strukturen – über bloße numerische Manipulation hinaus – ist.

Die Indiana University in Bloomington ehrte Karl Schuessler 2003 damit, dass das „Institute of Social Research“ in „Karl Schuessler Institute of Social Research“ umbenannt wurde.

Karl Schuessler war zwischen 1979 und 1990 mehrfach Gastprofessor sowohl an der Universität Mannheim als auch bei ZUMA. Mit seinen Workshops hat er die Forschungsaktivitäten des Instituts und seiner Mitarbeiter maßgeblich gefördert. So war er an der Vorbereitung und Durchführung einer Umfrage des in den 1980er Jahren noch existierenden Sozialwissenschaftenbusses beteiligt, deren Resultate in einer ZUMA-Publikation „Soziale Empfindungen. Ein interkultureller Skalenvergleich bei Deutschen und Amerikanern“ (Frankfurt, 1987) niedergelegt wurden. Er hat an den Geschicken des Instituts, das in den 1980er Jahren in den Dachverband der GESIS eingegliedert wurde, regen Anteil genommen. In diesem Zusammenhang sei auf den „ZUMA-Battle-Song“ verwiesen, den Karl Schuessler in den damaligen Turbulenzen für ZUMA geschrieben hat. Mannheim war für Karl Schuessler ein bisschen alte Heimat, und er erzählte gerne von seinen Vorfahren aus Alpirsbach im Schwarzwald.

Neben der Empirischen Sozialforschung hatte die Musik einen zentralen Stellenwert für Karl Schuessler. Karls Instrument war die Klarinette und seine Jazzcombo war ihm fast genauso wichtig wie seine Professur. Mit großem Stolz zeigte er seinen Mitgliedsausweis der Musikergewerkschaft. Über Jahre hinweg hat er auf den Jahrestagungen der American Sociological Association zusammen mit anderen (wie etwa George Bohrnstedt) die Teilnehmer mit seinem Dixieland unterhalten. Und so durfte dann auch bei dem Symposium anlässlich seines 75. Geburtstags in der Reimers-Stiftung in Bad Homburg im Jahre 1990

ein kleines Jazzkonzert mit Karl und George nicht fehlen (die Symposiums-Beiträge sind in einer Festschrift mit dem Titel: „New Directions in Attitude Measurement“, Berlin, 1993 zusammengefasst worden).

Karl Schuessler war als Mensch und Wissenschaftler eine herausragende Persönlichkeit. Seine Bescheidenheit, sein trockener Humor und sein Witz gepaart mit der Schärfe und Exaktheit wissenschaftlicher Analyse machten ihn zu einem wunderbaren, unvergesslichen Freund.

Wir schätzen uns glücklich, dass wir diese Freundschaft erleben durften.

Giessen, New York, im Mai 2006

DAGMAR KREBS & MANFRED KUECHLER

NACHRUF AUF PHILIP JAMES STONE

6. Oktober 1936 bis 31. Januar 2006



“Dear Faculty and Graduate Students,

It is with deep regret that I must inform you of Phil Stone’s sudden and unexpected death last night. Phil had been very sick but his passing comes as a surprise to all of us. The family will have a private service this weekend. A Memorial Service for Phil will be planned and details will follow later.”

Die Nachricht, die mich am 1. Februar 2006 erreichte, war kurz und ließ viele Fragen offen, die auch beim Memorial Service in der Memorial Church der Harvard University nicht beantwortet wurden.

Philip James Stone Ph.D. ‘60, Professor of Psychology, October 6, 1936 - January 31, 2006 lauten die nüchternen Daten auf der „Order of Service“ am 17. März 2006 in der Harvard University Memorial Church. Verdiente Kollegen, langjährige Freunde und Studenten zeichneten dort ein Bild mit einem Facettenreichtum wissenschaftlicher Verdienste und menschlicher Größe, die jeder, der Phil getroffen hat, spürte, die er selbst aber nie demonstrativ ausgebreitet hat.

Philip Stone studierte an der Universität Chicago im „great books program“, das er mit einer herausragenden Arbeit über „Soziale Wahrnehmung“ abschloss, die im Journal of Abnormal and Social Psychology veröffentlicht wurde. Dann studierte er an der Harvard Graduate School bei Gordon Allport und R. F. Bales. Die Forschungsarbeiten für seine Dissertation konzentrierten sich am NIMH und Walter Reed Hospital auf „physiological changes during competitive social interaction“. Gerade noch Graduate Student und Junior Faculty erhielt er – exzeptionell für Harvard – schon Tenure. Seine Wirkungsstätte als Harvard Professor fand er in der 14. Etage der William James Hall.

Die Forschungsarbeiten und seine Lehrveranstaltungen galten Themen wie group dynamics, computer-assisted content analysis, cross-national surveys of time use, cognitive

induction and narrative memory, information management, environmental psychology, and organizational behavior.

Seine Bücher, die er meist mit Kollegen oder Studenten herausgab, umfassen: *Experiments in Induction*; *The General Inquirer: A Computer Approach to Content Analysis*; *The Analysis of Communication Content*; *The Use of Time*; *Computer Recognition of English Word Senses* und *The SYMLOG Practitioner*. Neben seiner Tätigkeit im Psychology Department war er im Steering Committee für das gemeinsame Ph.D. Program in Organizational Behavior mit der Harvard Business School.

Im Vordergrund der Würdigung seines wissenschaftlichen Werkes stehen Pionierleistungen im Bereich der Inhaltsanalyse, der Zeitbudgetforschung und der Erkenntnisse der Netzwerkanalyse auf Businessarchitektur Design.

Phil Stone entwickelte mit Erwin K. Scheuch, der in den 60er Jahren selbst in Boston lehrte, die Konzepte zur Übertragung inhaltsanalytischer Verfahren auf die Informationsererschließung für Fragen aus Umfragen der Sozialforschung. Schon bald kam Phil Stone für ein Jahr nach Köln. Mit Hilfe des International Social Science Council wurde eine große internationale Konferenz über Inhaltsanalyse in Pisa bei CNUCE 1974 vorbereitet, die viele junge Wissenschaftler endgültig für die Weiterentwicklung und Anwendung neuer computerunterstützter Verfahren der Inhaltsanalyse begeisterte. Großes Vorbild wurde und bleibt der von Stone und Kollegen in den 60er Jahren entwickelte „General Inquirer“ mit dem Harvard Dictionary, das in späteren Versionen auch leistungsfähige Disambiguierungsroutinen enthält. Phil hob aber auch Verfahren empirischer Klassifikation als wichtige komplementäre Verfahren hervor. Seine Anwendungen konzentrierten sich auf die Analyse von Themen in Texten, einschließlich der Analyse von Antworten auf offene Fragen oder von Mediatrends.

Als die Daten der ersten internationalen Zeitbudgetuntersuchung (Szalai et al.) sich in mehreren Lochkartenschränken mangels leistungsfähiger Großrechner in Europa stauten und mit Fachzähl-Sortiermaschinen oder universitären Stecktafel-programmierten Rechnern angesichts der Informationsfülle nicht sinnvoll auswerten ließen, half Stone mit Unterstützung des Harvard Computing Centers aus. Er wandte sich einem Thema zu, das heute weitaus populärer ist als damals „On being up against the Wall: Women’s Time Patterns in Eleven Countries“.

Dann entdeckte Phil mit Robert Lucchetti vom Visual Studies Department ein neues Forschungs- und Entwicklungsfeld. Im Bereich der Umweltpsychologie stützt er sich auf den „activity settings approach“ von Roger Barker. Exemplarisch für diese Entwicklungen steht seine Schilderung für die bedarfsgerechte Gestaltung moderner Büros im Computer-

zeitalter. Seine Analysen zeigen den Bedarfswandel von großen einzeln genutzten Räumen hin zu dem Bull desk room, der zur sozialen Vermittlung der Mitarbeiter in Besprechungen dient, umgeben von vielen kleinern Räumen mit PC-Arbeitsplätzen. Noch heute sind seine Konzepte in diesem Bereich gefragt. Diese Arbeiten wurden vom französischen Minister für Kultur ausgezeichnet.

Beeindruckend sind die überwältigenden persönlichen Würdigungen. Allen gemeinsam ist eine zentrale Beobachtung: Er verfügte über die große Gabe, auf andere Menschen zuzugehen und selbst die scheuesten unter ihnen so anzusprechen, dass ihr Selbstbewusstsein geweckt wurde und sie ihre besten Talente entfalten konnten. Mit dem Tod von Phil Stone verlieren die Sozialwissenschaften eine überragende intellektuelle Ressource, viele von uns einen sehr guten Freund.

Köln, im Mai 2006

EKKEHARD MOCHMANN

DER ROLLING CROSS-SECTION SURVEY – EIN INSTRUMENT ZUR ANALYSE DYNAMISCHER PROZESSE DER EINSTELLUNGSENTWICKLUNG

Bericht zur ersten deutschen RCS-Studie anlässlich der
Bundestagswahl 2005

THE ROLLING CROSS-SECTION SURVEY – AN INSTRUMENT FOR ANALYZING THE DYNAMICS OF PUBLIC OPINION

A Documentation of the First German RCS Survey Conducted in the
Context of the 2005 National Election

RÜDIGER SCHMITT-BECK, THORSTEN FAAS & CHRISTIAN HOLST

Der Beitrag stellt eine neue, Panel-Befragungen komplementäre Methode zur Analyse dynamischer Prozesse der Meinungskristallisation und des Meinungswandels vor: den *Rolling Cross-Section Survey*. Die Grundidee des RCS-Designs besteht darin, die Durchführung einer auf einer Zufallsstichprobe basierenden Querschnittsbefragung in strikt kontrollierter Weise so über einen vorab definierten Zeitraum zu spreizen, dass die Befragten jedes einzelnen Tages jeweils in sich eine Zufallsstichprobe aus der Grundgesamtheit konstituieren. Solche Umfragen ermöglichen in höchst flexibler Weise Analysen von dynamischen Phänomenen der öffentlichen Meinung. Der Beitrag erörtert detailliert die idealtypischen Eigenschaften von RCS-Studien und diskutiert Probleme ihrer praktischen Umsetzung am Beispiel der ersten deutschen RCS-Studie, die während des Wahlkampfes zur Bundestagswahl 2005 realisiert wurde. Er schließt mit einigen exemplarischen Analysen, welche Rückschlüsse auf die Datenqualität ermöglichen und die besonderen Vorzüge von RCS-Daten verdeutlichen.

The present paper describes a new methodological approach for analysing dynamic processes of opinion crystallization and change: the *Rolling Cross-Section Survey*. The basic idea of the RCS design is to spread a regular cross-section survey that is based on a random sample over an a priori defined period of time in such a way that not only the entire sample, but also the respondents interviewed on any given day of the field period

constitute random samples. Data generated from RCS surveys give researchers unique opportunities to study dynamic phenomena of public opinion in highly flexible ways. Following an overview of the ideal-typical features of the RCS design, the paper discusses problems of implementing these principles in practical fieldwork, taking the example of the first German RCS study that was conducted during the campaign of the German National Parliamentary Election 2005. The paper closes with exemplary analyses that indicate the quality of these data and illustrate their usefulness.

1 Einleitung

Bevölkerungsumfragen haben in den letzten Jahrzehnten sowohl in der politikwissenschaftlichen Forschung als auch in der medialen Beobachtung des politischen Prozesses weite Verbreitung gefunden. In beiden Bereichen stellen sie heute so etwas dar wie den „Goldstandard“ (Brady 2000: 47) für die Messung der Orientierungen der Bürger zum politischen System, zu politischen Akteuren, politischen Problemen und den diskutierten Lösungsvorschlägen. Aus politikwissenschaftlicher Perspektive bilden Umfragen ein ausgezeichnetes Instrument, um zu testen, wie sich Theorien in der Konfrontation mit empirischen Beobachtungen bewähren. Gerade das Ziel, Theorien mittels Umfragen zu prüfen, war immer wieder ein Ansporn für methodische Innovationen. Computer-unterstützte Telefoninterviews, aber auch Möglichkeiten der Internet-basierten Befragung stellen nur zwei Beispiele dar, wie technische Neuerungen neue Möglichkeiten für die Umfrageforschung eröffnen.

Eine dieser Innovationen der Umfrageforschung steht im Mittelpunkt des vorliegenden Beitrages: das so genannte „*Rolling Cross-Section*“-Design (im Folgenden: RCS). Das zentrale Ziel von RCS-Studien besteht darin, dynamische Prozesse der Kristallisierung und Veränderung von Einstellungen und anderen Orientierungen abzubilden und zu analysieren. Anwendungen finden sich bisher ausschließlich im Bereich der Wahlkampfforschung. Im Prinzip stellt das RCS-Design aber auch für andere Phänomene kurzfristigen Wandels der öffentlichen Meinung einen attraktiven Analyseansatz dar. Immer nämlich besteht das Erfordernis, über einen längeren Zeitraum hinweg die öffentliche Meinung zu beobachten. Die Struktur des befragten Personenkreises muss dabei allerdings im Zeitverlauf konstant bleiben; ansonsten wäre es nicht eindeutig möglich, beobachtete Änderungen im Antwortverhalten als tatsächliche Änderungen der öffentlichen Meinung zu interpretieren, da sie ebenso Folge einer veränderten Struktur des befragten Personenkreises sein könnten. Dies ist allerdings bei herkömmlichen Umfragen der Fall: Stichproben gehen en bloc ins Feld, einige Zielpersonen antworten früher, andere später. Unterscheiden sich dann aber früh von spät gegebenen Antworten, könnte dies auf veränderte Rah-

menbedingungen oder aber auf strukturell verschiedene Antwortmuster hindeuten. Nur im ersten Fall wäre die Diagnose einer veränderten öffentlichen Meinung real; im zweiten Fall wäre sie ein Artefakt. An dieser Stelle setzt die Grundidee von RCS-Erhebungen an: Sie besteht darin, die Durchführung einer auf einer Zufallsstichprobe basierenden Querschnittsbefragung in strikt kontrollierter Weise so über einen vorab definierten Zeitraum zu spreizen, dass nicht nur alle Befragten zusammen, *sondern auch die Befragten jedes einzelnen Tages* jeweils in sich eine Zufallsstichprobe aus der Grundgesamtheit konstituieren.

Anlässlich der Bundestagswahl 2005 wurde die erste Umfrage dieses Typs in Deutschland realisiert. Ziel dieses Projektes ist die Analyse der dynamischen Wirkungen des Wahlkampfes zur vorgezogenen Bundestagswahl 2005 auf die Orientierungen der Wähler und darüber vermittelt auf das Wahlergebnis (Schmitt-Beck 2005). Der vorliegende Beitrag schildert die in diesem Projekt gemachten Erfahrungen. Idealtypisch betrachtet von bestechender Eleganz und Erklärungskraft, konfrontiert das RCS-Design den Forscher in der Praxis mit erheblichen Umsetzungsproblemen, auf welche die Literatur bislang kaum eingegangen ist. Wir werden im Folgenden zunächst idealtypisch beschreiben, worin die besonderen Eigenschaften von RCS-Studien bestehen. Anschließend wird erläutert, wie dieses anspruchsvolle Design in unserer Studie praktisch realisiert wurde. Dabei wird auch diskutiert, welche Aspekte sich als besonders schwierig erwiesen haben. Der Beitrag schließt mit einigen exemplarischen Analysen, welche Rückschlüsse auf die Datenqualität ermöglichen und die besonderen Vorzüge von RCS-Daten verdeutlichen.

2 Wahlkampfforschung und der *Rolling Cross-Section Survey*

Die Analyse der Wirkungen von Wahlkämpfen ist ein Spezialisierungsgebiet der Politikwissenschaft, das lange ein Schattendasein fristete und erst in den letzten zehn Jahren einen deutlichen Aufschwung genommen hat (Holbrook 1996; Farrell & Schmitt-Beck 2002; Brady & Johnston 2006; für einen aktuellen Überblick siehe Schoen 2005: 521-541). Das hat mit der lange vorherrschenden Vorstellung zu tun, dass Wahlkämpfe keinen wesentlichen Beitrag zur Erklärung von Wahlergebnissen leisten könnten, aber auch mit den besonderen methodischen Herausforderungen dieses Forschungsfeldes (Brady u. a. 2006).

Wahlkämpfe können als Ketten von Ereignissen konzeptualisiert werden (Hillygus & Jackman 2003: 584; Stöss 1997: 104). Diese werden überwiegend von Parteien und Kandidaten initiiert, können aber auch andere Hintergründe haben, etwa Publikationsentscheidungen von Medien, Handlungen anderer politischer Akteure (z. B. Interessengruppen), Entwicklungen der internationalen Politik oder sogar zunächst unpolitische Vorkommnisse wie z. B. Naturkatastrophen. Alle diese Ereignisse können sich auf die Entwicklung der Unterstützung von Parteien und Kandidaten auswirken, sofern sie Informationen transpor-

tieren, die zwischen diesen in wertender Weise diskriminieren und für die breite Öffentlichkeit wahrnehmbar werden. Dies wird in erster Linie durch die Vermittlung der Massenmedien erreicht (Holbrook 1996; Shaw 2000). Aufgabe der Wahlkampfforschung muss es demnach sein, Veränderungen von Einstellungen, Wahrnehmungen und Verhaltensabsichten sichtbar zu machen und sie mit Ereignissen des Wahlkampfes als möglichen Ursachen in Verbindung zu bringen. So können die Hintergründe der Dynamik der öffentlichen Meinung im Verlauf von Wahlkämpfen erklärt werden (Holbrook 1996: 48). Studien zur Analyse von Wahlkampfwirkungen müssen also längsschnittlich angelegt werden. „Because campaigns are dynamic phenomena, good campaign studies must be dynamic, too. Time must enter the analysis either directly or indirectly (as proxy for campaign activities and events).“ (Bartels 2006: 2; zit. n. d. Manuskript)

Einer der bedeutendsten methodischen Fortschritte in diesem Bereich war die Entwicklung des Panel-Designs mehrfacher Wiederholungsbefragungen derselben Personen durch Paul Lazarsfeld und seine Mitarbeiter Mitte des vergangenen Jahrhunderts. Die bahnbrechende Erie County-Studie zum Präsidentschaftswahlkampf 1940, bei der eine Stichprobe der Einwohner eines Landkreises in Ohio siebenmal in monatlichen Abständen befragt wurde (Lazarsfeld u. a. 1944), stellt eine bis zum heutigen Tag eindrucksvolle Pionierleistung dar. Panel-Studien sind seither die gebräuchlichste Methode, um dynamische Prozesse kurzfristiger Meinungsentwicklungen einzufangen. Panel-Designs haben den großen Vorteil, Vorgänge individuellen Wandels nachvollziehbar zu machen. Allerdings weisen sie auch einige gravierende Nachteile auf. Dazu zählen Konditionierungseffekte, aber auch die unvermeidliche Panel-Mortalität. Für Wahlkampfforscher ist insbesondere die zeitliche „Grobkörnigkeit“ solcher Erhebungen ein Problem. Aus logistischen Gründen ist es schwierig, zahlreiche Panel-Wellen in sehr dichter Folge zu realisieren.¹ Je mehr Zeit aber zwischen den einzelnen Panel-Wellen verstreicht, desto schwieriger ist es, Panel-Effekte eindeutig einem bestimmten Ereignis zuzuschreiben, weil die Zahl der potenziell für Effekte verantwortlichen Ereignisse mit der Zeit steigt. Außerdem weist das Instrument hinsichtlich möglicher Untersuchungsgegenstände eine gravierende Einschränkung auf. Der typische Untersuchungsplan, um Wirkungen eines Ereignisses mittels Panel-Daten zu analysieren, ist das Pre-Post-Design, d. h. die Durchführung von Befragungen vor und nach einem mutmaßlich einflussreichen Ereignis. Das ist aber nur möglich, wenn vor Studienbeginn schon entschieden wird, welches Ereignis untersucht werden soll. Nicht antizipierbare Vorkommnisse sind mit diesem Instrument ebenso wenig analysierbar

1 Mit der Verbreitung neuer Medien und dem Aufbau von Pools befragungswilliger Personen bieten sich allerdings auch in diesem Bereich neue Möglichkeiten, die sich etwa das OmniNet von FORSA zunutze macht. Im Vorfeld der Bundestagswahl 2002 wurde damit ein elf-welliges Panel realisiert (vgl. Güllner et al. 2005).

wie Ereignisse, die zwar im Prinzip antizipierbar sein mögen, für die aber mangels einer geeigneten Theorie keine Erwartungen politischer Effekte existieren. Für die in einer solchen Situation erforderlichen explorativen Analysen sind reine Panel-Designs ungeeignet (Johnston & Brady 2002; Romer u. a. 2004; Brady & Johnston 2006; Brady u. a. 2006; Bartels 1999, 2006).

Mit dem *Rolling Cross-Section-Design* wurde in den letzten Jahren eine neue Methode zur Analyse zeitlichen Wandels von Einstellungen und anderen Orientierungen entwickelt und konzeptionell perfektioniert, welche diese Probleme nicht aufweist (Johnston 2001; Johnston & Brady 2002; Brady & Johnston 2006; Brady u. a. 2006; Romer u. a. 2004). Amerikanische Kampagnenberater setzen Verfahren mit ähnlicher Logik schon länger (unter der Bezeichnung „tracking polls“) ein, um ihren Klienten möglichst detaillierte Daten über die Resonanz ihrer Wahlkampfmaßnahmen zu verschaffen (Blakeman 1995; Asher 1998: 112-3). Dies setzt eine *kontinuierliche* repräsentative Beobachtung der Zielgruppe voraus, was herkömmliche Umfragen in der Regel nicht leisten. Dort nämlich werden bestimmte Zielpersonen eher zu Beginn der Feldzeit, andere dagegen erst zu einem späteren Zeitpunkt erreicht und interviewt. Idealerweise aber müssen die Befragten eines jeden Tages jeweils in sich eine Zufallsstichprobe der Grundgesamtheit darstellen. Genau dies leisten RCS-Studien. Um die Beobachtung der öffentlichen Meinung im Zeitverlauf noch weiter zu standardisieren, wird darüber hinaus versucht, jeden Tag ungefähr gleich viele Interviews zu realisieren. Die für die inferenzstatistische Absicherung von Befunden zwingend notwendige Zufallsauswahl von Befragten kommt bei RCS-Studien auf doppelte Weise zum Tragen: Erstens im Hinblick auf die Chance eines Mitglieds der Grundgesamtheit, überhaupt in die Studie einbezogen zu werden, zweitens im Hinblick auf den Zeitpunkt innerhalb des Erhebungszeitraumes, an dem eine ausgewählte Befragungsperson interviewt wird. Zentrale Parameter, um diese distinktiven Merkmale von RCS-Studien zu gewährleisten, sind die Stichprobenziehung und die Feldadministration:

- Die Stichprobenziehung (präziser: die Ziehung der Haushaltsstichprobe) erfolgt bei einer RCS-Studie in zwei Stufen. Zunächst wird eine normale Zufallsstichprobe gezogen. Diese wird dann jedoch nicht wie üblich sofort komplett oder in unsystematischer – oftmals rein feldlogistischen Erwägungen folgender – Weise über einen gewissen Zeitraum freigegeben. Vielmehr wird sie in einem zweiten Schritt *zufällig* in Teilstichproben (so genannte *Replikate*) unterteilt, deren Anzahl der Zahl der geplanten Erhebungstage entspricht. Jedes Replikat stellt ebenfalls eine zufällige Stichprobe aus der Grundgesamtheit dar. Diese werden dann nach einem festgelegten „Fahrplan“ jeweils an einem bestimmten, zufällig ausgewählten Tag für die Befragung freigegeben.

- Nach der Freigabe wird für jedes Replikat eine Bearbeitungsroutine in Gang gesetzt, die einem strengen Protokoll folgt. Diese hat das Ziel, durch intensive Feldarbeit jede Teilstichprobe möglichst gut auszuschöpfen. Dies ist notwendig, weil in der Feldpraxis nicht alle Mitglieder eines Replikats umgehend am Tag der Freigabe interviewt werden können. Das wird nur für leicht erreichbare (und befragungsbereite) Personen der Fall sein, nicht jedoch für schwerer erreichbare Personen. Mit diesen kommen Interviews typischerweise erst nach mehreren vergeblichen Kontaktversuchen zustande. Essentiell für das RCS-Design ist nun, dass dieses Protokoll für alle Replikate in strikt einheitlicher Weise abgearbeitet wird. Um die Gleichheit aller Replikate im Hinblick auf ihre Repräsentativität für die Grundgesamtheit und die dafür notwendige Eigenschaft der zufälligen Auswahl zu gewährleisten, muss jedes von ihnen nach exakt denselben Kontaktierungsregeln bearbeitet werden. Weder der Start- oder Wochentag noch sonstige Gesichtspunkte dürfen dabei eine Rolle spielen. Diese Regeln beinhalten einen genau festgelegten, für alle Replikate gleich langen, mehrtägigen Zeitraum, für den gemeinsame Adressen bzw. Telefonnummern eines Replikats aktiv bleiben und Kontaktversuche unternommen werden. Auch für letztere gilt im Hinblick auf Anzahl und Zeitpunkte eine präzise Routine. Würden diese Abläufe geändert, würden nicht mehr alle Mitglieder der Grundgesamtheit mit derselben Wahrscheinlichkeit in die Stichprobe inkludiert und einem bestimmten Replikat alloziert. Damit könnte die für das RCS-Design zentrale Annahme der völligen strukturellen Gleichartigkeit aller Replikate nicht mehr aufrechterhalten werden.²

Die dichte zeitliche Taktung von RCS-Studien auf Tagesbasis und die Notwendigkeit strikter Kontrolle über den Befragungszeitpunkt präjudiziert den Erhebungsmodus: tägliche RCS-Erhebungen können nur telefonisch durchgeführt werden.³ Bei genauer Umsetzung der beschriebenen methodischen Anforderungen erhält man eine Umfrage, die alle Eigenschaften einer normalen Querschnittsstudie auf der Grundlage einer Zufallsstichprobe besitzt und auch als solche analysiert werden kann. Zugleich aber kann sie unter zeitli-

2 Dies impliziert, dass die zeitliche „Schrittgröße“ der Erhebungen während der gesamten Feldzeit nicht variiert werden darf. RCS-Studien müssen nicht zwingend auf täglicher Basis erfolgen, denkbar sind beispielsweise auch Erhebungen im Wochenrhythmus. Allerdings ist zu bedenken, dass größere Abstände zwischen Replikaten auf Kosten der Genauigkeit bei der Analyse der Effekte von Ereignissen gehen, weil dann eine präzise Ursachenattribution zu ganz bestimmten Ereignissen erschwert wird. Völlig unvereinbar mit der Logik des RCS-Design sind aber ungleichmäßige oder auch systematisch variiierende Abstände zwischen den Erhebungszeitpunkten. Ebenso ausgeschlossen ist die – bei Standardbefragungen übliche – Beendigung der Befragung nach Erreichen einer bestimmten Fallzahl.

3 Es gibt zwar persönlich-mündlich durchgeführte RCS-Studien, beispielsweise die Niederländische Wahlstudie 1998, doch arbeitete diese mit wöchentlicher Taktung.

chen Aspekten in beliebiger und völlig flexibler Weise in ebenfalls zufällige Unterstichproben für beliebige Zeitpunkte oder -perioden zerlegt werden. Durch diese Eigenschaft von RCS-Daten werden höchst differenzierte Analysen des dynamischen Geschehens im Verlauf von Wahlkämpfen möglich. Die Replikate, die bei der Erhebung von RCS-Daten zentrale Bedeutung haben, treten bei diesen Analysen nicht mehr in Erscheinung. Die Auswertungen orientieren sich vielmehr an den Kalendertagen, an denen die Interviews realisiert werden. Auf ihrer Grundlage können „synthetische Querschnittsstichproben“ gebildet werden. Diese beinhalten alle Interviews, die innerhalb eines vom Forscher definierten Zeitraums durchgeführt wurden.

Abbildung 1 **Zeitliche Struktur eines Rolling Cross-Section Survey**

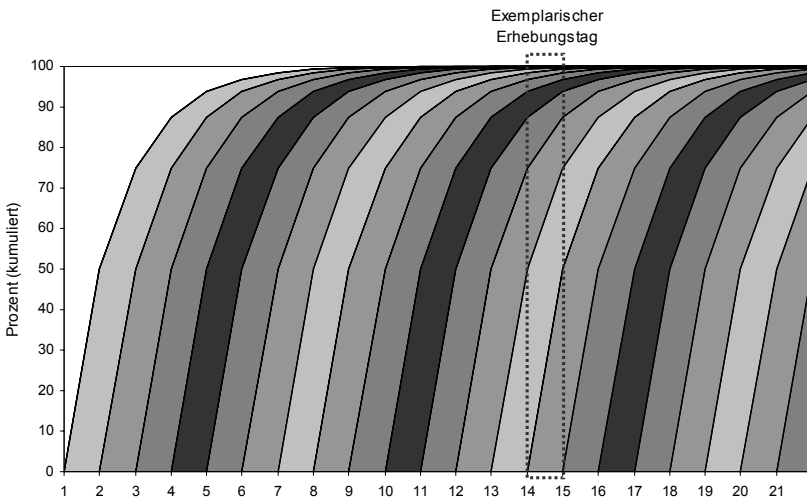


Abbildung 1 verdeutlicht die idealtypische Grundstruktur der durch RCS-Erhebungen gewonnenen Daten.⁴ Die Logik des RCS-Designs gewährleistet, dass die für jeden Erhebungstag zur Verfügung stehenden Interviews jeweils identische Mischungen von Befragten aus mehreren Replikaten und damit von leicht, aber auch schwer erreichbaren Personen beinhalten. Mit Blick auf die Stichprobe insgesamt ist der Befragungstag innerhalb

4 Bei der Darstellung handelt es sich um eine simulierte Datenstruktur unter der Maßgabe, dass jeden Tag das zu diesem Zeitpunkt in jedem Replikate noch vorhandene Potenzial noch nicht realisierter Interviews zu jeweils 50 Prozent ausgeschöpft wird.

der Feldzeit damit nicht mehr systematisch, sondern rein zufällig bedingt und besitzt somit den Charakter eines Zufallsereignisses. Die an den verschiedenen Erhebungstagen realisierten Interviews unterscheiden sich durch nichts außer dem Datum ihrer Durchführung. Noch präziser formuliert unterscheiden sie sich nur hinsichtlich des situativen Kontextes im Hinblick auf das Wahlkampfgeschehen, das diesem Datum entspricht. Alle Änderungen von Wahrnehmungen, Einstellungen, Präferenzen oder Verhaltensorientierungen, die im Zeitverlauf identifiziert werden, können daher nur zwei Quellen haben: den zufälligen Stichprobenfehler und realen Wandel unter dem Eindruck des Kampagnengeschehens. Diese können durch geeignete statistische Verfahren separiert werden (Johnston & Brady 2002; Brady & Johnston 2006; Romer u. a. 2004).

Auf diesem Wege gewonnene Daten können in äußerst flexibler und vielfältiger Weise analysiert werden. Man kann die Daten auf der Basis einzelner Erhebungstage analysieren und damit punktgenaue Verbindungen zwischen Wahlkampfereignissen und zugehörigen Änderungen der öffentlichen Meinung herstellen.⁵ Auch die Dauerhaftigkeit solcher Veränderungen ist leicht zu beobachten.

Neben der tageweisen Betrachtung können die Daten ebenso nach Formalkriterien zusammengefasst werden, beispielsweise wochenweise, um zu höheren Fallzahlen und damit höherer Teststärke zu gelangen. Da die Daten gleichmäßig über den gesamten Untersuchungszeitraum verteilt sind, können Schnittpunkte aber auch in beliebiger Weise nach inhaltlichen Kriterien gesetzt werden, etwa in Form eines Vergleiches der Interviews unmittelbar vor und unmittelbar nach einem bestimmten Ereignis im Wahlkampf. Ein besonderer Vorzug ist dabei, dass dies keineswegs nur für antizipierbare Ereignisse wie z. B. Parteitage oder TV-Duelle gilt, sondern für jedes beliebige Ereignis, einschließlich solcher, die im Vorfeld überhaupt nicht vorhergesehen werden konnten. Letztlich ist jede beliebige Kumulation von Daten aufeinanderfolgender Erhebungstage möglich. „From the sampling perspective, all that adding or subtracting a day does is reduce or expand standard error. [...] So the design combines flexibility in periodization with power in combination of days.” (Brady & Johnston 2006: 12; zit. n. d. Manuskript)⁶

5 Dass sich über das Datum als Schlüssel externe Daten problemlos zuspäuspielen lassen, beispielsweise aus Inhaltsanalysen der tagesaktuellen Medienberichterstattung, erleichtert auch die inhaltliche Interpretation beobachteter Änderungen (Gidengil & Dobrzynska 2003; Dobrzynska u. a. 2003; Fournier u. a. 2004).

6 Gelegentlich findet man in der Literatur auch Analysen normaler Querschnittsdaten, welche das Erhebungsdatum als unabhängige Variable einsetzen; allerdings ist das methodisch fragwürdig, weil bei solchen Studien der Befragungszeitpunkt mit der Erreichbarkeit der Befragten konfundiert ist, die ihrerseits häufig mit inhaltlich relevanten Merkmalen korreliert.

Nicht übersehen werden kann allerdings, dass RCS-Umfragen im Vergleich zu Panel-Befragungen auch einen deutlichen Nachteil aufweisen: Sie erlauben keine Aussagen über individuellen Wandel, sondern nur Aggregataussagen über Prozesse des Wandels in der gesamten Wählerschaft oder in Wählergruppen. Von daher bietet es sich an, RCS-Studien durch eine Panel-Komponente anzureichern, welche genau dies ermöglicht. Panel- und RCS-Studien können als komplementär gelten und ergänzen einander. Die Kombination einer RCS-Erhebung als Vorwahlbefragung mit einer als Nachwahlbefragung konzipierten zweiten Panel-Welle hat sich in den letzten Jahren als ideales Verfahren erwiesen, um die Kampagnendynamik bei Wahlkämpfen zu analysieren (Johnston 2001; Brady & Johnston 2006; Bartels 2006). Auch die hier beschriebene RCS-Erhebung wurde durch eine als Nachwahlbefragung realisierte zweite Panel-Welle angereichert. Die Variablen der Nachwahl-Welle können als Messungen des kumulativen Einflusses des gesamten Wahlkampfes interpretiert und als solche mit unterschiedlich weit in der Wahlkampfperiode zurückliegenden Vorwahl-Messungen derselben Variablen verglichen werden. Überdies lassen sich verschiedene für die Wahlkampfdynamik wesentliche Daten erst nach der Wahl sinnvoll erheben, nicht zuletzt natürlich die tatsächliche Wahlentscheidung selbst.

3 Die RCS-Studie zur Bundestagswahl 2005

3.1 Grundkonzept

Die erste RCS-Studie wurde als Pilotprojekt im Rahmen der American National Election Study (ANES) 1984 realisiert (Bartels 1988; Allsop & Weisberg 1988; Whiteley 1988). Allerdings erfolgte die Datenerhebung hier noch in wöchentlichem Turnus. Der eigentliche Durchbruch des RCS-Designs setzte die Umstellung auf tägliche Interviews voraus, welche erstmals im Rahmen der Canadian Election Study (CES) 1988 vorgenommen wurde. Seitdem ist das (mit einer Nachwahl-Panel-Welle kombinierte) RCS-Design bei der CES Standard (Johnston u. a. 1992, 1996; Nevitte u. a. 2000; Blais u. a. 2002). 1996 übernahm die nationale Wahlstudie Neuseelands (NZES) dasselbe Modell (Vowles u. a. 1998, 2002, 2004), seit 2001 beinhaltet auch die britische nationale Wahlstudie (BES) eine RCS-Komponente (Sanders et al. 2006; Clarke u. a. 2004, 2006). Die niederländische nationale Wahlstudie 1998 enthielt ebenfalls eine RCS-Komponente, jedoch mit wöchentlichem Erhebungsturnus (Aarts u. a. 1999). Eine RCS-Befragung konstituiert auch den Kern des größten jemals realisierten Forschungsprogramms der Wahlsoziologie: Der anlässlich der US-Präsidentenwahlen 2000 und 2004 durchgeführte, wahrhaft monumentale National Annenberg Election Survey (NAES) umfasst bis zu 300 tägliche Interviews über einen Zeitraum von 14 Monaten hinweg (Romer u. a. 2004; Johnston u. a. 2004; Clymer 2006). Schließlich wurden auch für die Bundestagswahlen 1994 und 2002

Analysen vorgelegt, die auf einem RCS-Umfragen ähnlichen Design basieren, den seit 1991 von FORSA durchgeführten kommerziellen OmniTel-Tagesbefragungen (Stöss 1997; Erhardt 1998; Dörmer 1998; Güllner u. a. 2005). Diese haben allerdings den Nachteil eines uneinheitlichen Erhebungsrhythmus: Nur an Werktagen werden Befragungen durchgeführt (Güllner 2000), was den idealtypischen Anforderungen eines RCS-Designs widerspricht. Ob es noch andere Abweichungen gibt, kann nicht beurteilt werden, da die genauen Details der Feldarbeit nach unserer Kenntnis nicht wissenschaftsöffentlich dokumentiert sind.

Im Vorfeld der vorgezogenen Bundestagswahl 2005 wurde die erste RCS-Studie in Reinform als Wahlstudie in Deutschland realisiert. Ziel des Projektes ist die Analyse der dynamischen Wirkungen des Wahlkampfes zu dieser ungewöhnlichen Wahl auf die Orientierungen der Wähler und darüber vermittelt auf das Wahlergebnis (Schmitt-Beck 2005). Im Folgenden beschreiben wir, wie das idealtypisch äußerst elegante, praktisch aber ungewöhnlich anspruchsvolle RCS-Design bei dieser Studie umgesetzt wurde und welche Erfahrungen wir dabei gemacht haben.⁷

Die RCS-Studie zur Bundestagswahl 2005 wurde in der Zeit vom 8. August bis zum 17. September 2005 durchgeführt; sie umfasste also einen Zeitraum von sechs Wochen und endete am letzten Tag vor der Bundestagswahl. Sie wurde als CATI-Erhebung realisiert und umfasst 3.583 zufällig ausgewählte Befragte aus der Grundgesamtheit der deutschsprachigen, in Privathaushalten mit mindestens einem Festnetzanschluss lebenden Bevölkerung der Bundesrepublik Deutschland, die zur Bundestagswahl 2005 wahlberechtigt war. Die Feldarbeit oblag dem Institut *Ipsos GmbH* (Möln). Die Stichprobenziehung erfolgte als ADM-Telefonstichprobe gemäß dem Gabler/Häder-Modell (Gabler & Häder 1997, Häder & Gabler 1998). Da eine Nachziehung von Telefonnummern designbedingt ausgeschlossen werden musste, war es notwendig, eine sehr umfangreiche Bruttostichprobe generierter Telefonnummern zu ziehen. Diese wurde dann in zufällige Teilstichproben zerlegt. Um ausreichend Reservetelefonnummern verfügbar zu haben, wurden insgesamt 60 Replikate zu je 600 Telefonnummern gezogen. Damit stand eine Bruttostichprobe von insgesamt 36.000 generierten Telefonnummern zur Verfügung, von denen letztlich 25.000 genutzt wurden.⁸ Angestrebt war zunächst, täglich 80 Interviews zu realisieren; nach 11 Tagen wurde diese Zielmenge auf 90 Interviews pro Tag erhöht.

7 Die hier beschriebene Studie wurde in engem Kontakt mit den Primärforschern von CES, NAES bzw. BES Richard Johnston, Elisabeth Gidengil, David Northrup und Harold Clarke geplant, denen wir für wertvolle Informationen und Anregungen zu großem Dank verpflichtet sind.

8 Diese sehr umfangreiche Bruttostichprobe wurde aus zwei Gründen für notwendig erachtet: Es war nicht bekannt, welches Brutto nötig war, um die angestrebten Interviewzahlen pro Tag zu erreichen, außerdem war zum Zeitpunkt der Planung der Studie noch nicht klar, ob die Bundes-

In unseren Vorüberlegungen hatten wir antizipiert, dass die Ausschöpfungsquoten nicht an allen Wochentagen gleich sein würden und möglicherweise auch in Abhängigkeit von nicht antizipierbaren Faktoren schwanken könnten. Es war daher notwendig, bei der Stichprobenziehung Voraussetzungen zu schaffen, die es erlauben würden, die Teilstichproben kontinuierlich mit dem Ziel möglichst gleichmäßiger täglicher Fallzahlen genauer auszusteuern. Dabei durfte allerdings die Logik des RCS-Designs nicht aufgegeben werden. Die dafür entwickelte Lösung sah vor, die Teilstichproben – wiederum zufällig – noch weiter in „Scheiben“ zu je 100 Telefonnummern zu zerlegen. Auch diese Scheiben waren aufgrund ihrer zufälligen Ziehung beliebig kumulierbar; es handelte sich gleichsam um „Replikate 2. Ordnung“. Dadurch wurde die Möglichkeit geschaffen, die Teilstichproben in einer dem RCS-Design angemessenen Weise ad hoc durch Hinzuspielen oder Weglassen einzelner Scheiben zu vergrößern oder zu verkleinern, um so die täglichen Fallzahlen realisierter Interviews möglichst konstant zu halten. Die Festlegung der Zahl der an jedem Erhebungstag neu einzuspielenden Scheiben erfolgte jeweils zeitnah in Reaktion auf die aktuelle Feldentwicklung und die bis dahin gesammelten Erfahrungen mit dem Verlauf der Erhebung. Alle Teilstichproben und Scheiben erhielten individuelle Kennungen; für jeden einzelnen Fall im Datensatz können daher Interviewzeitpunkt, Teilstichprobe bzw. Replikat sowie Scheibe innerhalb dieser Teilstichprobe eindeutig bestimmt werden.

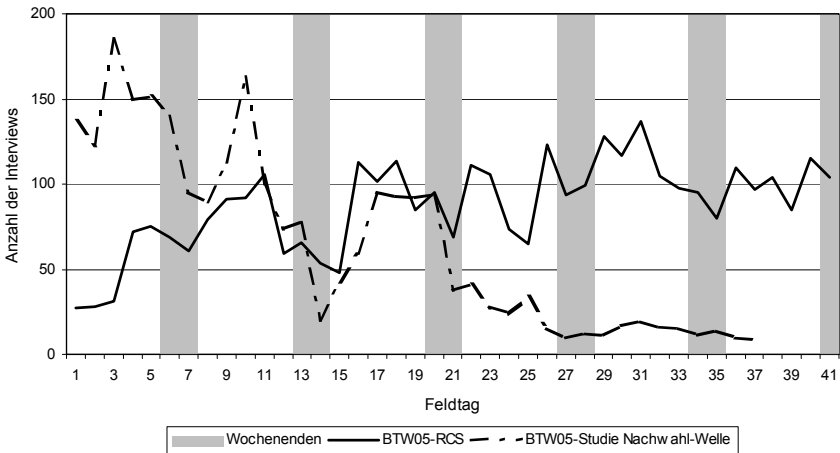
3.2 Feldverlauf

RCS-Studien können in der Erhebungsphase keinesfalls „sich selbst überlassen“ werden. Die Feldentwicklung bedarf permanenter Beobachtung und gegebenenfalls zeitnaher Adjustierung, sonst besteht ein erhebliches Risiko, das Ziel eines gleichmäßigen Feldverlaufs auf dem angestrebten Niveau täglicher Fallzahlen zu verfehlen. Faktisch wurde im Feldverlauf der Bundestagswahlstudie fast immer mit allen 600 Nummern, d. h. 6 Scheiben gearbeitet. An einem Tag wurde aufgrund sehr guter Feldentwicklung versuchsweise auf eine Scheibe verzichtet. Diese Maßnahme führte zu einer deutlichen Verringerung der Fallzahl und wurde daher am Folgetag wieder zurückgenommen. An einigen anderen Tagen wurden zusätzliche (7.) Scheiben eingesetzt, um absehbar geringere Fallzahlen zu kompensieren. Zumeist betraf dies Wochenenden.

tagswahl tatsächlich am 18. September 2005 stattfinden würde, weil noch die Entscheidungen des Bundespräsidenten und des Bundesverfassungsgerichts ausstanden. Die Zahl der Teilstichproben wurde in Anbetracht der nicht ausschließbaren Möglichkeit eines späteren Wahltermins bewusst sehr hoch angesetzt.

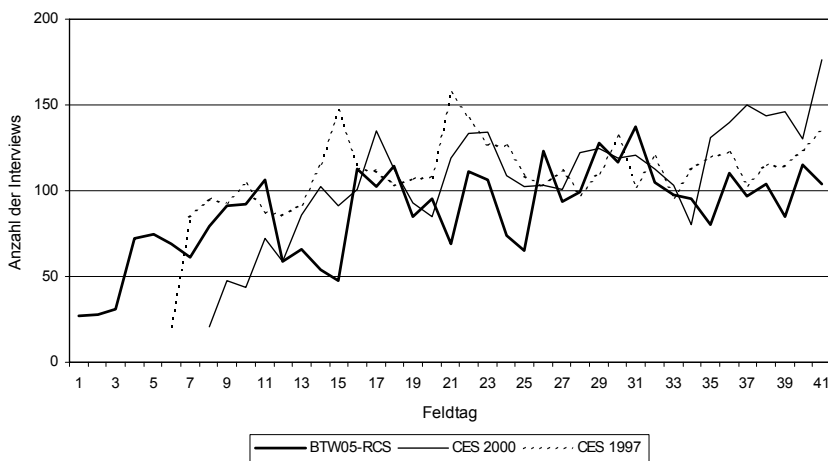
Wie sich der Feldverlauf der Bundestagswahlstudie entwickelte, ist Abbildung 2 zu entnehmen. Sie gibt die Zahl der Interviews wieder, die an jedem Feldtag realisiert wurden. Zum Vergleich ist auch der Feldverlauf der Nachwahl-Welle unseres Projektes eingezeichnet, der exemplarisch für die Entwicklung einer typischen Querschnittsbefragung stehen kann. Hier war die Maxime, möglichst viele Interviews in möglichst kurzer Zeit zu realisieren. Die Feldentwicklung der Nachwahlstudie verlief zwar nicht völlig monoton, insbesondere sind Einbrüche an den Wochenenden, und da wiederum vor allem sonntags, deutlich zu erkennen. Aber der Gesamttrend entwickelte sich klar von sehr hohen Zahlen realisierter Interviews in den ersten Tagen der Feldperiode (die sofort am Montag nach der Bundestagswahl einsetzte) zu geringeren täglichen Ausbeuten und schließlich einer Endphase mit schleppendem Feldverlauf auf niedrigem Niveau. Im Vergleich dazu stellt sich, als Folge der systematischen Zeitsteuerung, der Feldverlauf der RCS-Welle sehr viel gleichmäßiger dar.

Abbildung 2 Zahl der täglichen Interviews: RCS-Studie zur Bundestagswahl 2005 im Vergleich zur Nachwahl-Welle



Quelle: RCS-Studie „Kampagnendynamik 2005“

Abbildung 3 Zahl der täglichen Interviews: RCS-Studie zur Bundestagswahl 2005 im Vergleich zur CES 1997 und 2000



Quelle: RCS-Studie „Kampagnedynamik 2005“, Canadian Election Studies 1997 und 2000 (vgl. Northrup 1998: 7, 2003: 11)

Gemessen am Ziel einer möglichst gleichmäßigen Zahl täglicher Interviews mag der Feldverlauf indessen immer noch recht unregelmäßig erscheinen. So wird erkennbar, dass die ersten Tage deutlich hinter den Zielzahlen zurückblieben. Das ist zum Teil eine logische Konsequenz der Stichprobenallokation, denn die angestrebten täglichen Fallzahlen können erst dann erreicht werden, wenn eine größere Zahl von Replikaten gleichzeitig aktiv ist, was unmittelbar nach Feldbeginn natürlich noch nicht der Fall sein konnte. Ein gewisses „Aufschaukeln“ am Anfang ergibt sich zwangsläufig aus der Logik des RCS-Designs und zeigt sich analog auch bei vergleichbaren anderen Studien. Das belegt Abbildung 3 am Beispiel der kanadischen Wahlstudien 1997 und 2000. Eine wichtige Implikation hiervon ist im Übrigen, dass die Interviews der ersten Erhebungstage noch keine perfekten Zufallsstichproben darstellen und daher nicht in dynamische Analysen einbezogen werden können (Johnston & Brady 2002: 285). Das ist auch in der idealtypischen Darstellung von Abbildung 1 erkennbar. Im Fall der deutschen Studie kamen aber in den ersten Feldtagen Probleme mit dem Adressenverwaltungssystem hinzu, auf die weiter unten zurückzukommen ist.

Festzustellen ist aber auch ein insgesamt nicht völlig ebener Feldverlauf mit größeren Tagesschwankungen. Das betrifft zunächst einmal die Wochenenden, und da wiederum

vor allem die Sonntage, an denen – wie bei der Nachwählerhebung – stets weniger Interviews realisiert werden konnten als an den Werktagen zuvor und danach. Wie beschrieben, wurde dieser Tendenz teilweise durch den Einsatz zusätzlicher Scheiben an den Wochenenden entgegengewirkt. Das hat verhindert, dass diese Einbrüche noch größer wurden. Aber solche Interventionen sind nicht unbegrenzt möglich, weil sonst designbedingte „Aufschaukelungsprozesse“ an den Folgetagen ausgelöst würden, an denen diese Scheiben ja ebenfalls aktiv bleiben. Daher konnten die „Wochenendtäler“ aus systematischen Gründen nicht vollständig nivelliert werden. Durchaus erhebliche, aber nicht systematisch deutbare Unterschiede zeigen sich überdies auch zwischen den Werktagen.⁹ Das ist allerdings auch bei den beiden kanadischen Vergleichsstudien nicht anders.¹⁰ Die Fallzahlen täglich realisierter Interviews schwanken bei der Bundestagswahlstudie zwischen 27 am ersten Feldtag und 137 am 31. Feldtag bei einem Gesamtdurchschnitt von 87.4 Interviews pro Tag und einer Standardabweichung von 26.7. Erwähnung verdient auch ein Phänomen, das sich in allen kanadischen RCS-Studien zeigte (siehe auch Johnston & Brady 2002: 285), nicht jedoch in der deutschen: eine – vermutlich seitens der Befragten motivationsbedingte – Zunahme der Ausschöpfung der Replikate in den letzten Tagen vor der Wahl, die sich in einem deutlichen Anstieg der realisierten Fallzahlen niederschlug. Die angestrebten täglichen Fallzahlen erweisen sich somit in der Praxis als Richtgröße, während die Zahl faktisch realisierter Interviews offenkundig unvermeidbaren Fluktuationen unterworfen ist – ein Phänomen, das alle RCS-Studien betrifft. So präzise, dass ein völlig ebener Feldverlauf erzielt wird, lässt sich der Interviewprozess bei RCS-Studien nicht steuern, weil nur Instrumente zur Feinsteuerung eingesetzt werden können, welche die Erfordernisse des RCS-Designs nicht verletzen.

3.3 Bearbeitung der Replikate

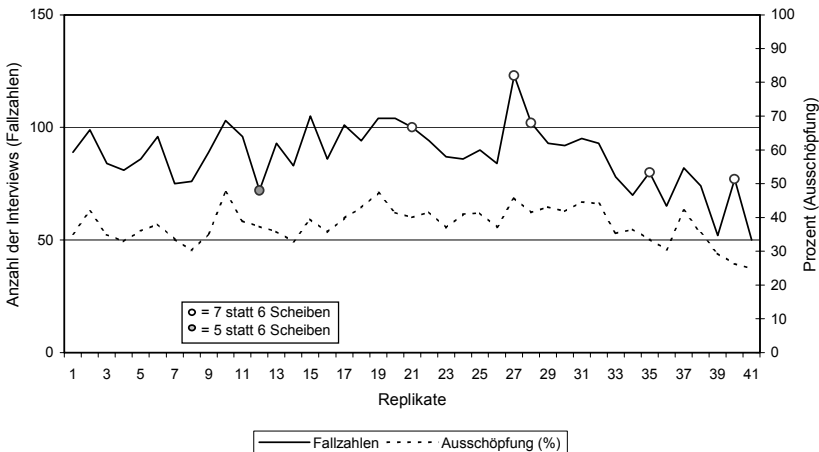
Neben der tageweisen Betrachtung ist auch eine Betrachtung der Replikate, die an jeweils einem Tag ins Feld gegeben und an diesem und den folgenden Tagen abgearbeitet wurden, möglich. Abbildung 4 zeigt die Zahl der Interviews, die aus jedem der 41 täglichen Replikate hervorgegangen sind. Zwischen den Teilstichproben sind deutlich weniger Schwan-

9 Ein Einbruch im Verlauf der Felddurchführung ist klar auf technische Probleme zurückführbar: Am 15. Erhebungstag konnten ab ca. 15:00 keine Interviews mehr durchgeführt werden, weil bei Bauarbeiten in der Umgebung des Telefonstudios ein Stromkabel durchtrennt und damit das gesamte Telefonstudio „lahm gelegt“ worden war – ein anschauliches Beispiel für die Unwägbarkeiten, welche die Umsetzung eines methodisch rigorosen Designs wie des hier beschriebenen beeinträchtigen können.

10 Für andere RCS-Studien sind leider keine numerischen Detailinformationen über täglich realisierte Fallzahlen dokumentiert; siehe jedoch auch die ganz ähnlichen Grafiken zur CES 1993 bei Johnston (2001: 155) und zur NAES 2000 bei Romer (2004: 58).

kungen zu verzeichnen als zwischen den Erhebungstagen (s. o.). Auffällig ist vor allem ein einsamer Gipfel für ein Replikat, aus dem 123 Interviews realisiert werden konnten. Zumindest teilweise ist das dadurch zu erklären, dass es sich hierbei um eine der Teilstichproben mit erhöhter Scheibenzahl, d. h. größerem Bruttoumfang, handelte. Auch einige der anderen „Gipfel“ sind auf erhöhte Scheibenzahlen zurückzuführen, eines der „Täler“ wurde korrespondierend durch Reduktion der Scheibenzahl bewirkt. Das standardisierte Bild der Ausschöpfungsquoten der Teilstichproben ist demgegenüber sehr viel gleichmäßiger. Auffällig ist die stetig schrumpfende Zahl der Interviews, die aus den letzten Teilstichproben realisiert wurden – ein Muster, das auch in anderen RCS-Studien festgestellt wurde. Dieses Phänomen ist designbedingt: Die letzten vor der Wahl ins Feld gegebenen Replikate konnten nicht für dieselbe Zeitdauer offen bleiben wie diejenigen, die so rechtzeitig ins Feld gegangen waren, dass sie noch vor der Wahl vollständig „abgearbeitet“ werden konnten. Da die Erhebung am Tag vor der Wahl eingestellt wurde „trunkierte“ der Wahltag diese Replikate und setzte ihrer Bearbeitung gleichsam ein natürliches Ende. Das ist idealtypisch auch aus Abbildung 1 ersichtlich. Für die späteren Analyse-möglichkeiten ist das jedoch unerheblich, denn die Stichprobenqualität der vorliegenden *Erhebungstage* – die letztlich entscheidend sind – wird dadurch nicht beeinträchtigt.

Abbildung 4 Realisierte Fallzahlen und Ausschöpfung aus Replikaten



Quelle: RCS-Studie „Kampagnendynamik 2005“

Die Daten der einzelnen Erhebungstage setzen sich jeweils aus Interviews zusammen, die sehr rasch nach Replikatöffnung durchgeführt wurden, und anderen, bei denen es länger, manchmal sogar sehr lange dauerte, bis sie realisiert werden konnten. Eine rigorose Feldbearbeitung mit dem Ziel, auch schwer erreichbare Befragte in die Studie einzubeziehen, ist kennzeichnend für RCS-Studien. Auch bei der deutschen Bundestagswahlstudie wurde auf eine außerordentlich intensive Feldarbeit Wert gelegt, um eine möglichst gute Ausschöpfung der einzelnen Replikate zu erreichen. Dazu gehörten beispielsweise gezielte Konversionsversuche „weicher“ Verweigerer¹¹: Sie wurden zwei Tage nach ihrer ursprünglichen Ablehnung noch einmal kontaktiert. Insgesamt 104 zusätzliche Interviews konnten so gewonnen werden, was einer Konversionsrate von 14.1 Prozent aus den ursprünglichen Verweigerungen entspricht. Die Interviewer konnten dabei, um ihre Argumentation zu unterstützen und glaubhaft zu machen, auf die Homepage von Ipsos verweisen, auf der eine Kurzbeschreibung des Projekts abgelegt war, sowie einen Link zu der ausführlicheren Projektdarstellung auf der Homepage der Primärforscher an der Universität Duisburg-Essen. Der Hinweis auf diese Informationsquelle konnte natürlich auch schon bei Befragten erfolgen, die in der ersten Kontaktpphase zögerlich waren.

Überdies wurden intensiv und über einen langen Zeitraum Kontaktversuche bei nicht erreichbaren, aber nicht als ungültig erkennbaren generierten Telefonnummern unternommen. Jedes Replikat blieb 21 Tage lang aktiv (um das Risiko zu minimieren, dass Kontaktversuche vollständig in Phasen urlaubsbedingter Abwesenheit möglicher Befragungspersonen fielen). In den ersten beiden Tagen sollte jede nicht sofort stichprobenneutral ausgeschiedene Telefonnummer, bei der noch kein Interview realisiert werden konnte, 8-mal angerufen werden, an den Folgetagen jeweils 2-mal. Im gesamten Feldverlauf nicht erreichbare Nummern waren somit insgesamt 54-mal anzurufen. Bezüglich der Uhrzeiten der Kontaktversuche war dabei ein komplexes Rotationsschema zu beachten. Telefoniert wurde jeden Tag von 10 bis 21 Uhr, sonntags von 14 bis 20 Uhr. Damit ging das Design der Bundestagswahlstudie noch einmal deutlich über die diesbezüglich auch schon sehr „aggressiven“ (Romer u. a. 2004: 59) anderen RCS-Studien hinaus.¹² Allerdings traten bei der Umsetzung dieses sehr aufwendigen Kontaktschemas Probleme auf. Diese hatten mit der Priorisierung der Kontaktversuche zu tun, die in den Algorithmus zur

11 Das sind Verweigerer, die z. B. unter der Begründung, keine Zeit oder kein Interesse zu haben, ein Interview ablehnten.

12 Das bei den CES angewandte Schema sah für jedes Replikat 2 bis 4 Kontaktversuche an den ersten vier Feldtagen und einen Kontaktversuch an den verbleibenden Feldtagen vor, und zwar für insgesamt 12 (CES 1988 und 1993) bzw. 10 Tage (CES 1997 und 2000). Bei der NAES 2000 erfolgten an den ersten vier Feldtagen je zwei Kontaktversuche, an den restlichen Tagen einer; jedes Replikat blieb 14 Tage aktiv (Romer u. a. 2004: 59).

Telefonnummernverwaltung einprogrammiert werden musste. Dabei erhielten Terminvereinbarungen die höchste Priorität¹³, gefolgt von besetzten Nummern (die nach 20 Minuten erneut zu kontaktieren waren), neu eingespielten Nummern aus dem aktuellen Replikat und schließlich den bereits eingesetzten Nummern aus den Replikaten der Vortage.

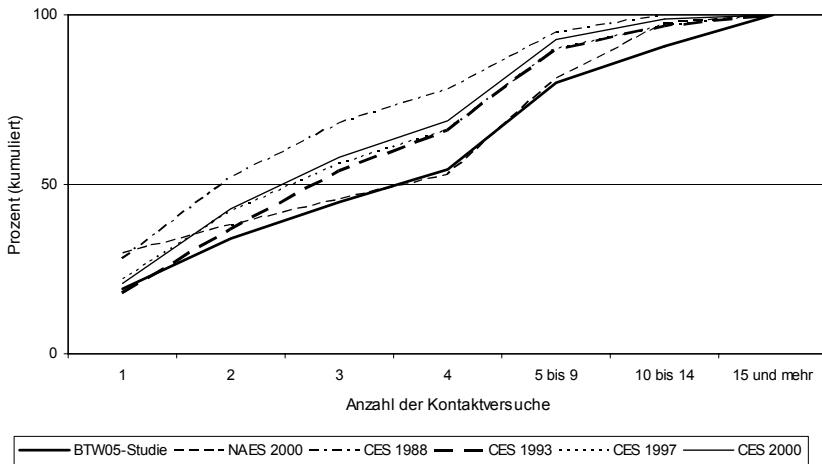
Vor dem Hintergrund der sehr großen Bruttostichprobe und der hohen Zahl von notwendigen Kontaktversuchen führte dieses Schema in der Praxis zu Klumpungen: Zu viele Nummern wurden zeitgleich ins System eingespielt, um sie alle tatsächlich kontaktieren zu können. In der Folge blieb die reale Zahl der Kontaktversuche unter dem angestrebten Soll. Als nach den ersten drei Feldtagen die realisierten Fallzahlen hinter den Erwartungen zurück blieben, wurde der Algorithmus geprüft und die Zeiten der Einspielung gleichmäßiger verteilt. In der Folge wurden dann die Sollfallzahlen rasch erreicht. Im Nachhinein zeigte jedoch eine aufwendige Analyse der Kontaktprotokolle, dass das angestrebte Kontaktschema auch danach nicht im vollen Umfang umgesetzt wurde. Dafür gab es während der Feldphase keine Indizien, weil es lediglich zu einer Verlangsamung des Feldverlaufes der einzelnen Teilstichproben führte, aber nicht zu einer sichtbaren Dämpfung der erreichten Fallzahlen. Wie sich bei der Nachanalyse zeigte, wurden die noch offenen Telefonnummern an den ersten beiden Öffnungstagen im Schnitt nur etwa halb so oft angerufen wie angestrebt (3.9- bzw. 3.8-mal statt 8-mal), an den Folgetagen (mit einem Soll von 2 Kontaktversuchen) lagen die durchschnittlichen Kontaktversuchsdaten zwischen 2.3 und 1.4. Hieraus ist eine Lehre für künftige Studien dieses Typs abzuleiten: Der Entwicklung und Programmierung des Adressverwaltungsalgorithmus, der die Priorisierung und Abarbeitung der Kontaktversuche steuert, muss im Vorfeld größte Beachtung geschenkt werden. Aufgrund seiner zwangsläufigen Komplexität wäre im Grunde ein Pretest wünschenswert. Doch würde das nichts anderes bedeuten, als dass die Studie einige Tage unter Feldbedingungen laufen müsste, weil sich Kontaktierungsprobleme nur dann zeigen können. Diese äußerst aufwändige und im Grunde prohibitiv kostspielige Option wurde von uns im Vorfeld nicht ins Auge gefasst; unser Projekt beinhaltete nur einen normalen Pretest mit wenigen Interviews, welcher der Optimierung des Fragebogens diente.

Abbildung 5 zeigt im Vergleich mit allen CES-Studien und der NAES 2000, wie viele Kontaktversuche für die Realisierung der Interviews nötig waren. Dabei wird deutlich, dass sich eine intensive Feldarbeit lohnt. In keiner RCS-Studie wurde mehr als etwa ein Viertel der insgesamt am Ende realisierten Interviews schon beim ersten Kontaktversuch erzielt. Lediglich die CES 1988 kam bereits beim zweiten Kontaktversuch über die 50-

13 Streng genommen verstoßen Terminvereinbarungen mit Befragungspersonen gegen die strikte Logik des RCS-Designs, aber sie sind bei Erhebungen nur um den Preis einer deutlich verschlechterten Ausschöpfung verzichtbar.

Prozent-Marke. Im Vergleich zu den anderen Studien erweist sich der Feldverlauf bei der deutschen Studie als besonders langwierig. Die 50-Prozent-Marke wurde erst nach vier Kontaktversuchen erreicht, nach 10 Kontaktversuchen lagen 80 Prozent der später insgesamt erzielten Interviews vor. Rund sieben Prozent der Interviews kamen erst nach mehr als 15 Kontaktversuchen zustande. In Einzelfällen wurden Interviews sogar erst nach mehr als 30 Kontaktversuchen realisiert.

Abbildung 5 Anzahl der Kontaktversuche bis zum Interview: RCS-Studie zur Bundestagswahl, NAES 2000 und CES 1988 - 2000



Quelle: RCS-Studie „Kampagnendynamik 2005“, National Annenberg Election Study 2000 (vgl. Romer u. a. 2004: 63; Email-Mitteilung von Richard Johnston am 19.01.2005), Canadian Election Studies 1988 - 2000 (vgl. Northrup 1989: 10, 1994: 16, 1998: 13, 2003: 11)

Tabelle 1 zeigt, von welchen Faktoren es abhing, ob ein Interview rasch oder erst nach zahlreichen vergeblichen Kontaktversuchen zustande kam. Aufgrund der erwähnten Trunkierung der verfügbaren Feldzeit für späte Replikate durch den Wahltermin ist klar, dass die Zahl der überhaupt für Interviews verfügbaren Tage eine wesentliche Rahmenbedingung darstellt, die kontrolliert werden muss, auch wenn sie selbst keinen substanziellen Erkenntniswert besitzt. Bei einer breiter angelegten explorativen Analyse konnten vier unabhängige Variablen identifiziert werden, die sich auf die Zahl der Kontaktversuche

auswirkten, die für ein Interview nötig waren; sie betreffen ausnahmslos persönliche Merkmale der Befragten.¹⁴ Eine höhere Anzahl von Kontaktversuchen war demnach nötig, um Interviews mit jüngeren Personen, mit Hochgebildeten, mit Vollzeitbeschäftigten und mit Personen in kleinen Haushalten, insbesondere Ein-Personen-Haushalten zu realisieren. Dies sind durchweg Merkmale, die eine geringere Erreichbarkeit für telefonische Kontaktversuche in Privathaushalten indizieren und bedingen, dass im Schnitt intensivere Kontaktanstrengungen erforderlich sind, um mit solchen Personen Befragungen durchzuführen.¹⁵

**Tabelle 1 Zahl der Kontaktversuche bis zum erfolgreichen Interview:
Determinanten**

	Standardisierter Regressions- koeffizient	t-Wert	p
Zahl der in Feldzeit verfügbaren Tage (1-21)	.27	17.23	.000
Alter (in Jahren)	-.09	-5.31	.000
Abitur (dummy)	.05	3.31	.001
Vollzeit beschäftigt (dummy)	.11	7.00	.000
Haushaltsgröße (Zahl der Personen)	-.04	-2.26	.024

N = 3.559

Modellgüte $R^2=10.7$ Prozent; dabei erklären die Kontrollvariablen eine zusätzliche Varianz von 3.1 Prozent gegenüber einem Modell, in das nur die Feldzeit eingeht.

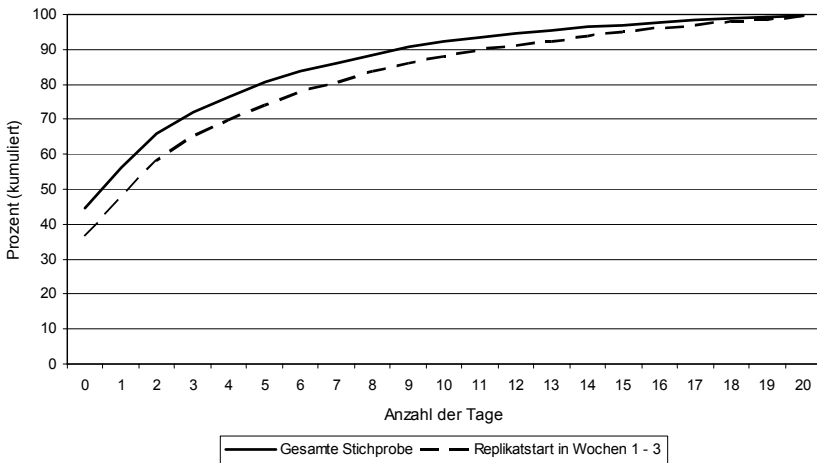
Daraus folgt unmittelbar, dass es mitunter sehr lange dauerte, bis Interviews realisiert werden konnten. Analog zu Abbildung 5 gibt Abbildung 6 wieder, wie viele Tage nach Replikatstart die vorhandenen Interviews durchgeführt wurden. Knapp 45 Prozent aller Interviews wurden ohne Verzögerung sofort am Tag der Freigabe ihres Replikats realisiert

14 Als nicht signifikante Prädiktoren der Zahl der Kontaktversuche bis zum erfolgreichen Interview erwiesen sich das Geschlecht der Befragten, ihre Kirchengangshäufigkeit, ihre Gewerkschaftsmitgliedschaft, die Größe ihres Wohnortes, die Wohnregion (alte oder neue Bundesländer), das politische Interesse, die Stärke der Parteibindung sowie die Zahl der Telefonanschlüsse im Haushalt und die Wiederbefragungsbereitschaft als persönliche Merkmale sowie der Wochentag des Replikatstarts als Merkmal der Erhebung.

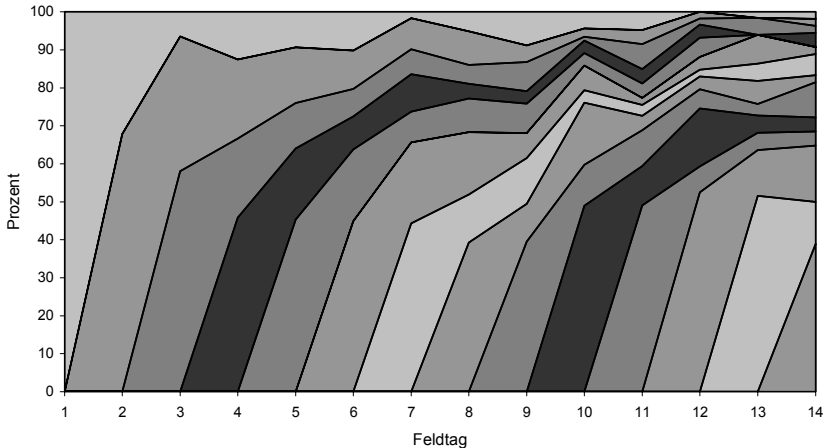
15 Die Zahl der für ein Interview der RCS-Studie nötigen Kontaktversuche ist positiv mit der Zahl der Kontaktversuche bei der 2. Welle und (leicht) negativ mit dem Vorliegen eines Interviews der 2. Welle korreliert (Pearsons Korrelationskoeffizient $r = .14$ bzw. $r = -.09$, jeweils signifikant mit $p < .001$ bei einem zweiseitigen Test).

(durchgezogene Linie), 56 Prozent innerhalb der ersten beiden Tage, an denen auch die intensivsten Kontakanstrengungen unternommen wurden. 84 Prozent aller vorliegenden Interviews entstammen der ersten Feldwoche, aus der dritten Feldwoche kommen knapp fünf Prozent. Dabei ist aber zu bedenken, dass dieser „Erfolgswert“ für die lange Feldzeit artifiziell gedämpft ist, denn schon nach der Hälfte der insgesamt sechswöchigen Erhebungsperiode der Studie gingen nur noch Replikate ins Feld, die wegen des herannahenden Wahltermins nicht mehr volle 21 Tage lang ausgeschöpft werden konnten. Das verdeutlicht der Vergleich mit der gestrichelten Linie, welche den Feldverlauf nur für diejenigen Replikate wiedergibt, die in den ersten drei Wochen freigegeben wurden und deshalb noch vor der Wahl vollständig abgearbeitet werden konnten. Von diesen Interviews stammen deutlich höhere Anteile aus späteren Phasen der jeweiligen Öffnungsperioden. Sechs Prozent wurden noch in der dritten Woche realisiert, über 22 Prozent in der zweiten Woche. Lediglich ein gutes Drittel dieser Befragungen wurde sofort am ersten Tag durchgeführt, die 50-Prozent-Marke wurde erst am dritten Tag erreicht.

Abbildung 6 Tage zwischen Replikatfreigabe und Interview



Quelle: RCS-Studie „Kampagnendynamik 2005“

Abbildung 7 **Anteile der Replikate an den täglichen Interviews**

Quelle: RCS-Studie „Kampagnendynamik 2005“

Abbildung 7 präsentiert eine Art Resümee der bisherigen Beobachtungen. Abbildung 1 gab einen Eindruck von der idealtypischen Datenstruktur einer RCS-Umfrage, diese Darstellung zeigt nun die tatsächliche Datenstruktur der in den ersten 14 Tagen der Erhebungsperiode unserer RCS-Studie realisierten Interviews. Wegen der Fluktuation der täglichen Fallzahlen basiert diese Darstellung im Interesse der Übersichtlichkeit allerdings auf relativen Häufigkeiten innerhalb einzelner Feldtage. Offenkundig war bereits am fünften Feldtag eine gute Mischung von Interviews aus mehreren Replikaten erreicht, deren dominante Struktur sich später nur noch wenig änderte. Zu fragen ist nun, ob diese Daten auch tatsächlich die Erwartungen erfüllen, denen RCS-Studien gerecht werden sollen.

4 Zur Qualität der RCS-Daten

4.1 Immanente Qualität

Wie steht es um die Qualität unserer Daten gemäß der immanenten Kriterien des RCS-Designs? Das Kernanliegen einer solchen Studie besteht darin, eine Stichprobe zu realisieren, für deren Mitglieder nicht nur die Inklusion selbst, sondern auch der Tag, an dem sie interviewt werden, per Zufall bestimmt wird. Wenn es gelungen ist, dies adäquat um-

zusetzen, dann kann davon ausgegangen werden, dass die Verteilungen demographischer Variablen und anderer stabiler Merkmalsdimensionen im Feldverlauf nicht variieren. Ob dies der Fall ist, wird im Folgenden getestet. Für RCS-Daten bieten sich in besonderer Weise graphische Verfahren der Datenanalyse an (Romer u. a. 2004; Brady & Johnston 2006; Clarke u. a. 2004, 2006). Trendplots von Skalenmittelwerten oder Anteilen interessierender Merkmale empfehlen sich stets als Einstieg, um einen ersten Eindruck von der dynamischen Entwicklung theoretisch interessanter Aspekte der öffentlichen Meinung im Erhebungszeitraum zu gewinnen. Graphische Analysen sind daher ein häufig benutztes Instrument in RCS-Projekten, das die formale Modellierung von Variablenzusammenhängen nicht nur vorbereitet, sondern auch sinnvoll ergänzt. Wenn man eine solche Analyse auf der Ebene der einzelnen Stichprobentage durchführen will, ist man allerdings mit dem Problem konfrontiert, dass die einzelnen Tagesstichproben jeweils relativ klein und daher zwangsläufig mit einem recht großen Zufallsfehler behaftet sind. Dem kann durch Verfahren der Datenglättung entgegengewirkt werden. Die meisten vorliegenden Analysen von RCS-Daten verwenden hierfür einfache Verfahren gleitender Durchschnitte. Anspruchsvollere und solchen Herangehensweisen überlegene Techniken der Punktschätzung mittels Verfahren der nicht-parametrischen Regression (Fox 2000; Elff 2006) wurden bislang nur selten eingesetzt (siehe jedoch Clarke u. a. 2004, 2006). Wir werden im Folgenden auf den Scatterplot-Smoother LOWESS zurückgreifen.¹⁶ Die Analysen beginnen jeweils am fünften Tag der Feldperiode; die in der Anlaufphase der Erhebung gewonnenen Interviews können – wie oben diskutiert – designbedingt nicht als Zufallsstichproben gelten und werden deshalb bei dieser Analyse nicht berücksichtigt.

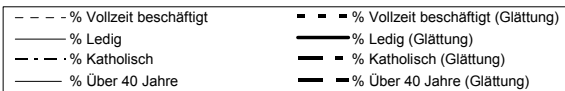
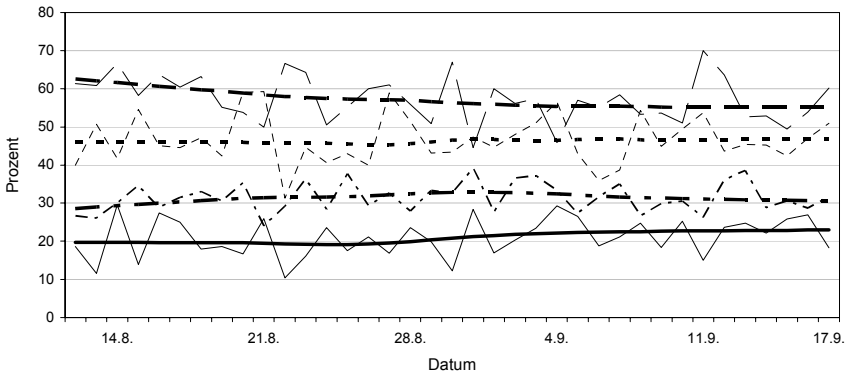
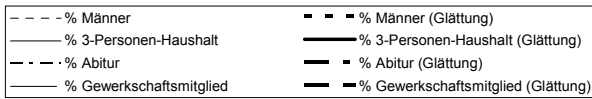
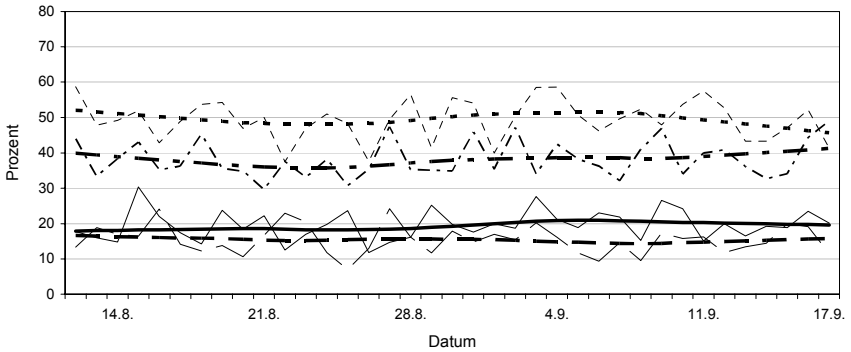
Wenn das Erhebungsverfahren das gewünschte Ergebnis – eine über Zeit völlig gleichmäßige Stichprobe – geliefert hat, bei der die Fälle jedes einzelnen Tages in sich eine Zufallsstichprobe darstellen, dann müssen unveränderliche demographische Merkmale in allen Tagesstichproben gleich verteilt sein. Systematische Trends dürfen nicht erkennbar

16 LOWESS steht für „*L*ocally *W*eighted Scatterplot Smoother“. Dieses Glättungsverfahren berechnet für jeden einzelnen Datenpunkt einer Punktwolke eine lokale Regression, in die nur die „nächsten Nachbarn“ dieses Punktes eingehen – im hier vorliegenden Fall für jeden Tag also die Werte der Befragten des jeweils gleichen sowie der benachbarten Tage. Für jeden Datenpunkt wird aus den beobachteten Punkten, welche diesen umgeben, ein Schätzwert ermittelt. Eine Gewichtungsfunktion sorgt dabei dafür, dass der Einfluss dieser Punkte mit zunehmender Distanz abnimmt; weiter entfernte Punkte determinieren dadurch den Schätzwert für einen Punkt schwächer als näher liegende. Dabei können verschiedene so genannte Bandbreiten gewählt werden. Allen in diesem Artikel präsentierten Analysen liegt eine Bandbreite von .5 zugrunde, d. h. für die Schätzung jedes Punktes wurde – mit abnehmender Gewichtung – die Hälfte aller Beobachtungen berücksichtigt. Die Verbindung dieser Schätzwerte ergibt am Ende die geglättete Trendlinie (vgl. Cleveland 1994: 168-180; Schnell 1994: 109-113; Fox 2000: 19-50).

werden, lediglich zufallsbedingte Fluktuationen sind zulässig. Demographische Merkmale dürfen also nicht mit den Zeitpunkten korreliert sein, zu denen Befragte interviewt wurden. Abbildung 8 belegt für einige exemplarisch ausgewählte Merkmale, dass dies in unserer Studie in der Tat weitgehend der Fall ist. Die Anteile der Männer, der Personen, die in Drei-Personen-Haushalten leben, der Abiturienten und der Gewerkschaftsmitglieder weisen zwar geringfügige Schwankungen um die jeweiligen Mittelwerte auf, d. h. die Feldverläufe waren nicht immer vollkommen eben, aber eine Tendenz nach oben oder unten lässt sich nicht ausmachen.¹⁷ Dasselbe gilt auch für die Anteile der Vollzeit-Erwerbstätigen und der Katholiken. Etwas anders verhält es sich bei den Ledigen – ihr Anteil scheint über die Zeit leicht zuzunehmen. Der Anteil der Befragten, die 40 Jahre und älter sind, nimmt demgegenüber leicht ab. Etwa drei Wochen nach Beobachtungsbeginn tritt dann jeweils eine Stabilisierung ein. Wir haben oben gesehen, dass Personen, die erst nach vielen Kontaktversuchen befragt werden konnten, im Schnitt jünger sowie häufiger voll erwerbstätig waren. Abbildung 8 deutet darauf hin, dass sich die Stichprobenstruktur bezüglich dieser Merkmale erst dann stabilisiert hat, als die ersten Replikate abgeschlossen wurden, so dass die Replikatekomposition der Tagesstichproben keinen weiteren Veränderungen mehr unterworfen war. Insgesamt sind aber auch diese Schwankungen gering.

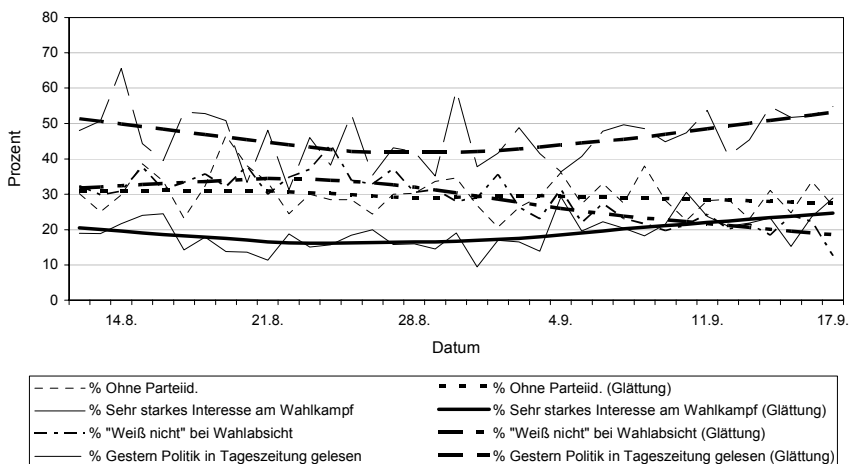
17 Eine inferenzstatistische Überprüfung bestätigt den Augenschein: Für alle möglichen Paare von Tagesstichproben – bei insgesamt 37 Tagesstichproben sind dies 666 – haben wir den Anteil der Männer und der Abiturienten mittels logistischer sowie das mittlere Alter mittels linearer Regressionen in den beiden jeweils betrachteten Tagesstichproben verglichen. In die Regressionsmodelle haben wir jeweils neben einer Konstanten nur eine einzige Dummy-Variable, die die Zugehörigkeit zu *einer* der beiden einbezogenen Tagesstichproben anzeigt, einbezogen. Für den Anteil der Männer erreichte der Koeffizient dieser Variable in 44 Fällen (6,6 Prozent) statistische Signifikanz ($p < .05$). Für den Anteil der Abiturienten bzw. das mittlere Alter waren es 54 (8,1 Prozent) bzw. 34 (5,1 Prozent). Da bei einem Signifikanzniveau von fünf Prozent 33 signifikante von insgesamt 666 Koeffizienten zu erwarten sind, deuten diese Werte auf fast ausschließlich zufällige Schwankungen der sozialstrukturellen Merkmale hin.

Abbildung 8 Verteilungsstabilität demographischer Merkmale



Quelle: RCS-Studie „Kampagnendynamik 2005“

Abbildung 9 Verteilungsentwicklung politischer Orientierungen



Quelle: RCS-Studie „Kampagnendynamik 2005“

Systematische Veränderungen der demographischen Zusammensetzung der Stichprobe von Erhebungstag zu Erhebungstag wären ein Hinweis auf mangelhafte Implementation des RCS-Designs. Abbildung 8 deutet darauf hin, dass solche Probleme nicht existieren. Anders verhält es sich bei den in Abbildung 9 wiedergegebenen politischen Orientierungen. Hier ist keine Stabilität über Zeit zu erwarten, jedenfalls dann nicht, wenn es sich nicht um tief in der Persönlichkeit verwurzelte Grundhaltungen handelt. Im Gegenteil – Wahrnehmungen und Einstellungen sollten sich ändern, denn die Erwartung, dass sich solche Orientierungen im Verlaufe von Wahlkämpfen wandeln, ist eine zentrale Prämisse und die Rechtfertigung der Durchführung von RCS-Studien. Solche Dynamiken sichtbar zu machen, ist ja gerade ihr Zweck. Abbildung 9 illustriert diese Überlegungen am Beispiel von vier unterschiedlichen politischen Orientierungen. Abgetragen ist zunächst der Anteil derjenigen Wähler, die sich nicht mit einer politischen Partei identifizieren. Die orthodoxe Theorie der Parteiidentifikation ging davon aus, dass solche Orientierungen außerordentlich stabil sind und sich unter normalen politischen Umständen kaum jemals ändern (Campbell u. a. 1960). Jüngere Deutungen halten Parteibindungen zwar für durchaus wandelbar; aber auch sie interpretieren sie als eher stabile Komponenten der politischen Persönlichkeit (Fiorina 1981; Schmitt-Beck & Weick 2001; Schmitt-Beck u. a. 2006). Abbildung 9 deutet darauf hin, dass sich der Anteil der parteipolitisch Unabhängi-

gen (und damit implizit auch der parteigebundenen Wähler) in der Tat während des Wahlkampfes zur Bundestagswahl 2005 praktisch nicht verändert hat. Insoweit bekräftigt dieser Befund den aus der Analyse der demographischen Variablen gezogenen Schluss zur immanenten Qualität der Daten.

Im Vergleich dazu waren die anderen drei politischen Orientierungen relativ beweglich. Sehr deutlich wird erkennbar, wie der Wahlkampf den Wählern geholfen hat, Präferenzen zu formieren. Der Anteil derjenigen Befragten, die zum jeweiligen Befragungszeitpunkt noch nicht wussten, wie sie wählen sollten, halbierte sich in den letzten vier Wochen vor der Wahl kontinuierlich von etwa 35 auf unter 20 Prozent. Im selben Zeitraum wuchs das Interesse am Wahlkampf. Korrespondierend dazu nahm der Anteil der Befragten zu, die angaben, sich am Vortag durch Lektüre einer Tageszeitung politisch informiert zu haben. Der nächste Abschnitt wird belegen, dass auch die Parteipräferenzen selbst deutlichen Veränderungsprozessen unterworfen waren.

4.2 Die Entwicklung von Parteipräferenzen: die RCS-Daten im Vergleich mit anderen Datenquellen

Im Mittelpunkt unseres Projektes stehen letztlich die Wahlabsichten und ihre Entwicklung während des Wahlkampfes bis hin zu ihrer letztendlichen Manifestation in Form des Ergebnisses der Bundestagswahl am 18. September 2005, das am Ende des Wahlkampfes verzeichnet wurde und zumindest partiell als Reaktion auf Prozesse der Präferenzbildung und -änderung in dessen Verlauf verstanden werden kann. Der RCS-Studie zufolge hat es im Verlauf des Wahlkampfes einige deutliche Verschiebungen in den Parteipräferenzen gegeben. Diese vergleichen wir im Folgenden mit anderen Messungen der Unterstützung der Parteien und ihrer Entwicklung. Naheliegender ist selbstverständlich zunächst der Vergleich mit dem Wahlergebnis selbst. Allerdings interessiert hier keineswegs nur die statische, auf den 18. September fokussierte Perspektive, sondern ebenso die dynamische Perspektive auf Entwicklungen im Vorfeld. Wie stellt sich diese Entwicklung der Unterstützung der Parteien über den gesamten Wahlkampf hinweg in den RCS-Daten dar?

Ein vergleichbares Anliegen verfolgen auch die regelmäßig zum Zwecke der Medienpublikation produzierten Versuche von Meinungsforschungsinstituten, Stimmenverteilungen für eine hypothetische Bundestagswahl zu einem bestimmten, vor der tatsächlichen Wahl liegenden Termin zu ermitteln. Die „demoskopischen“ Institute achten zwar normalerweise darauf, dabei den Begriff „Wahlprognose“ zu vermeiden, und weichen stattdessen auf weniger verfängliche Begriffe wie „Projektion“ oder einfach „Sonntagsfrage“ aus (vgl. ZDF 2006). Aber sie können dadurch kaum verhindern, dass die in vielen Medien publizierten Fieberkurven der Wählermeinung in der öffentlichen Wahrnehmung und Diskussi-

on häufig wie Vorhersagen des tatsächlichen Wahlergebnisses der kommenden Bundestagswahl verarbeitet werden und sich dadurch möglicherweise auch auf das Stimmverhalten der Wähler auswirken (Brettschneider 1992; Schmitt-Beck 1996). Das ist auch der Hintergrund der Debatte über die Qualität von Medienumfragen, welche – wie schon in vergleichbaren Situationen bei früheren Wahlen – sofort nach der Bundestagswahl 2005 einsetzte, weil die vor der Bundestagswahl publizierten Zahlen vom Wahlergebnis in eklatanter Weise abwichen (Hofmann-Göttig 2005). So schienen wieder einmal die „Demoskopken“ ein „Debakel“, wenn nicht gar ein „Desaster“ zu erleben und wurden als die „eigentlichen Wahlverlierer“ apostrophiert (vgl. Spiegel 2005, NZZ 2005, FTD 2005), während sich die SPD am Ende ihres Wahlkampfes als „Umfragesieger-Besieger“ (Das Parlament 2005) brüsten konnte.

Eine weitere Ausdrucksform der Entwicklung der öffentlichen Meinung gegenüber den Parteien im Verlauf des Wahlkampfes stellen die Kurse von Wahlbörsen dar. Eine Wahlbörse funktioniert wie ein Aktienmarkt – Aktien werden zu aktuellen Marktpreisen gekauft und verkauft. Im Falle von Wahlbörsen handelt es sich dabei um „Parteiaktien“. Da dem Besitzer einer Parteiaktie nach der Bundestagswahl ein Betrag in Höhe des Zweitstimmenergebnisses der jeweiligen Partei ausgezahlt wird, spiegeln die Kurse zu jedem Zeitpunkt die durchschnittlichen Erwartungen der Händler hinsichtlich des Wahlausgangs wider. Der aktuelle Kurs einer Aktie entspricht jeweils dem Preis, zu dem die letzte Transaktion getätigt wurde. Die Kurse können wie Stimmenanteile interpretiert werden. Anders als Befragungspersonen in Umfragen bringen die Teilnehmer von Wahlbörsen allerdings keine momentanen persönlichen Präferenzen zum Ausdruck, sondern ihre Vermutungen über den Ausgang der Wahl und damit letztlich über das voraussichtliche Verhalten aller anderen Wähler. Im Gegensatz zu Meinungsumfragen werden Wahlbörsen von ihren Betreibern daher als echte Prognoseinstrumente für das Wahlergebnis propagiert, die sich gleichsam der kollektiven „Weisheit des Marktes“ bedienen und dadurch sensibel auf aktuelle politische Entwicklungen reagieren (Brüggelambert 1999; Huber 2002). Wir berücksichtigen bei unserer Analyse Daten der *Wahlbörse* sowie der *Wahlstreet*.¹⁸ Um auf eine tageweise Basis zu kommen, betrachten wir die gemittelten täglichen Schlusskurse der beiden Börsen, d. h. für jede Partei den (mittleren) Preis der letzten Transaktion des Tages.

18 Beide Wahlbörsen wurden von ECCE TERRAM veranstaltet – die *Wahlbörse* im Auftrag von Financial Times Deutschland sowie T-Online, die *Wahlstreet* im Auftrag der ZEIT sowie des Handelsblatt (vgl. <http://wahlboerse.t-online.de>; <http://www.wahlstreet.de>). Unser Dank gilt Frank Simon von ECCE TERRAM, der uns die Daten freundlicherweise zur Verfügung gestellt hat.

Diese drei Datenquellen – RCS-Studie, Medienumfragen¹⁹ sowie Wahlbörsen – wollen wir daher mit Blick auf den 18. September selbst, aber auch mit Blick auf die Dynamik im Vorfeld des 18. September vergleichen. Anders als die in Abbildung 8 und 9 wiedergegebenen Befunde berücksichtigt die nachfolgende Analyse nur diejenigen Befragten unserer RCS-Studie, die angaben, zumindest „vielleicht“ an der Bundestagswahl teilnehmen zu wollen, und bei der Frage nach dem beabsichtigten Wahlverhalten eine Partei nannten.²⁰ Um die Zufallsschwankungen der tagesgenau gemessenen und damit nur auf relativ geringen Fallzahlen beruhenden Verteilungen zu dämpfen, werden die Daten wieder einer LOWESS-Glättung unterzogen.

Bekanntlich waren bei der Bundestagswahl 2005 alle Formen der Datensammlung mit einem schwierigen Umfeld konfrontiert, wie auch die einzelnen Grafiken in Abbildung 10 zeigen: Vergleicht man die Endpunkte der verschiedenen Kurven mit dem tatsächlichen Wahlergebnis vom 18. September 2005, so liefert letztlich keine Datenquelle eine exakte Entsprechung zum Wahlergebnis vom 18. September 2005 über alle Parteien hinweg. Insbesondere das schlechte Abschneiden der Union deutet sich in keiner Weise an.

Welches Bild zeichnen die verschiedenen Datenquellen aber für die Dynamik im Vorfeld der Wahl?²¹ Auffällig ist bei dieser Betrachtung zunächst, wie eng die von Medienumfragen und Wahlbörsen ermittelten Parteienteile stets beieinander lagen. Möglicherweise

19 Die Zeitreihen der Medienumfragen wurden wie folgt gebildet: Es wurden alle veröffentlichten „Sonntagsfragen“ bzw. „Projektionen“ von Emnid, Forsa, der Forschungsgruppe Wahlen, Infratest dimap sowie dem Institut für Demoskopie Allensbach im Zeitraum vom 13. August bis zum 18. September 2005 gesammelt (siehe auch <http://wahlrecht.de/umfragen/archiv/2005.htm>). Um daraus tageweise organisierte Zeitreihen zu bilden, wurden den einzelnen Tagen die Parteienteile aus der jeweils zu diesem Zeitpunkt aktuellsten Umfrage zugewiesen. Die Zeitreihen zeigen also unabhängig vom Institut für jeden Tag die publizierten Verteilungen der zuletzt veröffentlichten „Sonntagsfrage“ an. Wurden an einem Tag zwei neue Umfragen unterschiedlicher Institute veröffentlicht, wurden daraus Mittelwerte gebildet.

20 Für den Vergleich mit Medienumfragen und Wahlbörsen ist eine möglichst realitätsnahe Schätzung der täglichen (Rand-)Verteilungen dieser Frage wichtig. Da die RCS-Daten die bei Telefonumfragen oft zu beobachtende Überrepräsentation hoch gebildeter Befragter aufweisen (Glagow & Bühler 1987; Mosmann 1998), wurde bei den Berechnungen nach Bildung korrigiert. Das verwendete Gewicht passt die einzelnen Tagesstichproben an die Bildungsverteilung des ALLBUS 2004 an.

21 Diese vergleichende Darstellung kann dabei nicht im selben Sinne wie der Vergleich mit dem Wahlergebnis (im Sinne der Kriteriumsvalidierung) interpretiert werden, weil die Konstruktvalidität der „Projektionen“ abgesehen vom unmittelbaren zeitlichen Umfeld der Wahl eine unbekannte Größe ist. Es steht dann nämlich kein Außenkriterium als Vergleichsmaßstab zur Verfügung und die Institute legen die Verfahren, über welche sie ihre Rohdaten in „Projektionen“ transformieren, nicht offen (Wüst 2003). Ähnliches gilt auch für die Wahlbörsen. Der Vergleich unserer RCS-Studie mit diesen Daten stellt somit letztlich unterschiedliche Repräsentationen der öffentlichen Meinung einander gegenüber.

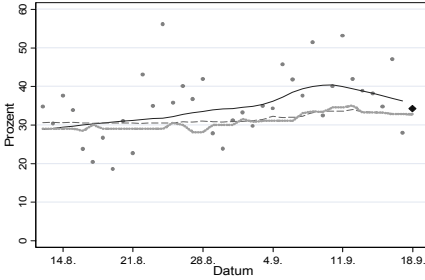
fließen publizierte Umfrageergebnisse als wesentliche Information in die Kaufentscheidungen der Händler an den Wahlbörsen ein, so dass deren Kurse faktisch gar keine unabhängigen Messungen darstellen.

Vergleicht man die geglätteten RCS-Werte über den gesamten hier betrachteten Zeitraum hinweg mit Werten aus Medienumfragen bzw. Wahlbörsen, zeigen sich viele Gemeinsamkeiten, aber auch einige interessante Unterschiede. Für die SPD etwa zeichnen sowohl die Medienumfragen als auch die Wahlbörsen zunächst ein träges Bild um die 30 Prozent; erst nach dem TV-Duell vom 4. September steigen die Werte der Sozialdemokraten bis auf 32 oder 33 Prozent an. Die RCS-Daten verzeichnen hingegen schon früher ein leichtes Ansteigen des SPD-Anteils, das sich nach dem TV-Duell deutlich beschleunigt. Überdies – und das ist das eigentlich Überraschende – legen die RCS-Daten den Schluss nahe, dass die Unterstützung der SPD zwar nach dem TV-Duell einen Höhepunkt erreichte, in den Tagen unmittelbar vor der Wahl aber wieder etwas schrumpfte. In der sport-affinen Sprache der Wahlkampfberichterstattung ausgedrückt, befand sich die SPD am Ende demnach gar nicht mehr in einer Aufholjagd – vielmehr ging ihr gegen Ende des Wahlkampfes eher schon wieder die Puste aus.

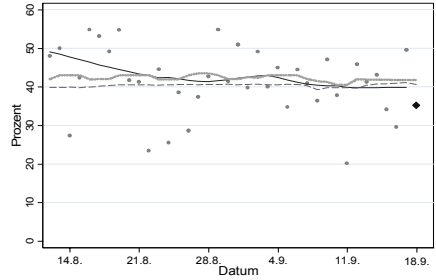
Für die Union stellt sich das Bild deutlich weniger bewegt dar, und zwar unabhängig von der Datenquelle. Sowohl für die RCS-Studie als auch für Medienumfragen und Wahlbörsen werden konstant Werte im Bereich um 40 Prozent für die Union registriert, die drei Zeitreihen überlappen sich stark. Den Niedergang der Union vor dem Wahltag, der schließlich zum Wahlergebnis von 35.2 Prozent führte, bildet keine der Datenreihen adäquat ab. Allerdings fallen die ungeglätteten Datenpunkte der RCS-Studie in den letzten Tagen vor der Wahl auffallend niedrig aus; lediglich am allerletzten Tag vor der Wahl wird der Stimmenanteil der Union – möglicherweise zufallsbedingt – wieder klar überschätzt. Bei der Glättung führen diese beiden gegensätzlichen Entwicklungen zu einer Nivellierung, aus der dann kein Trend mehr ablesbar ist. An den Schwierigkeiten, solche sich mutmaßlich sehr schnell vollziehenden und im Volumen nicht übermäßig starken, in den machtpolitischen Konsequenzen aber entscheidenden Bewegungen kurz vor der Wahl adäquat abzubilden, zeigen sich designbedingte Grenzen von RCS-Studien, die allerdings nicht systematischer, sondern forschungspragmatischer Natur sind. Mit mehr Befragten in den letzten Tagen der Erhebung hätte sich der Zufallsfehler reduzieren lassen und es wären genauere Messungen zustande gekommen – allerdings nur um den Preis einer sehr deutlichen Steigerung der Erhebungskosten.

Abbildung 10 **Entwicklungen der Parteistärken im Vorfeld der Bundestagswahl 2005: RCS-Studie im Vergleich zu Medioumfragen und Wahlbörsen**

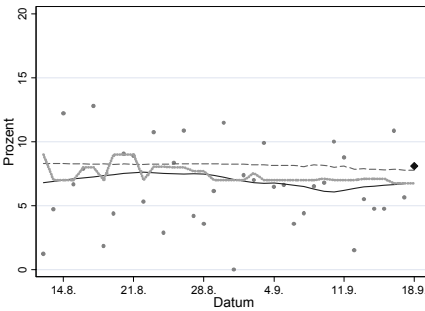
a) SPD



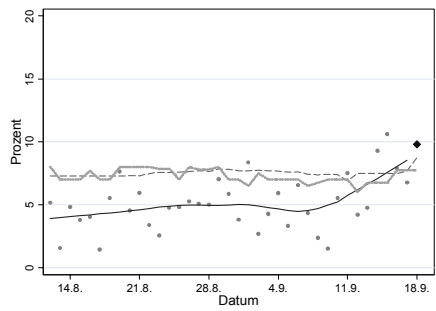
b) Union



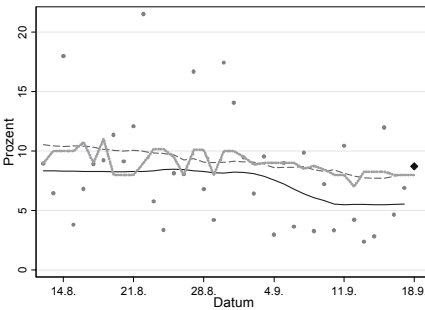
c) Grüne



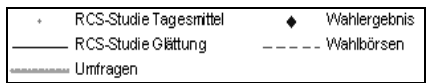
d) FDP



e) Linkspartei



Legende



Quelle: RCS-Studie „Kampagnendynamik 2005“, publizierte Umfragen von EMNID, Forsa, der Forschungsgruppe Wahlen, Infratest dimap und des Instituts für Demoskopie Allensbach sowie tägliche Schlusskurse der Internet-Wahlbörsen „Wahlstreet“ und „Wahlbörse“

Die Grünen liegen in den Wahlbörsen konstant bei etwa acht Prozent, ähnlich auch in den Medienumfragen. Die RCS-Daten deuten hingegen auf etwas stärkere Schwankungen hin und zeigen gerade am Wahlkampftage (leicht) wachsenden Zuspruch für die Grünen. Möglicherweise ist das ein Effekt strategischen Wahlverhaltens in Reaktion auf die in der letzten Phase des Wahlkampfes intensivierte, auf rot-grün-orientierte SPD-Anhänger zielende Zweitstimmenkampagne der Grünen. Dies könnte auch den gleichzeitigen leichten Schwund bei den SPD-Anteilen plausibel machen. Besonders interessant ist der Befund für die FDP. Wiederum zeichnen Wahlbörsen und Medienumfragen ein ziemlich konstantes Bild. An den RCS-Daten stechen jedoch zwei Dinge klar hervor: dieser Studie zufolge lag die FDP lange Zeit nur im Bereich um fünf Prozent (und damit deutlich unter den übrigen Werten des Prognosemarktes), um dann aber einen fulminanten Schlusspurt hinzulegen: In der letzten Wahlkampfwoche erst steigt die FDP auf ihr annähernd zweistelliges Wahlergebnis, was die LOWESS-Anpassungskurve auch klar sichtbar macht.²² Auch hierbei hat es sich vermutlich um ein Phänomen strategischen Wählens gehandelt, dieses Mal jedoch im bürgerlichen Lager. Die Werte der Linkspartei haben sich im Verlauf des Wahlkampfes kontinuierlich verschlechtert – diesen Schluss legen *alle* Trendlinien nahe. Dieses Absinken wird in den RCS-Daten allerdings noch markanter (und letztlich wohl zu markant) sichtbar als in Medienumfragen und Wahlbörsen.

Insgesamt legt diese vergleichende Analyse den Schluss nahe, dass das „Debakel der Demoskopen“ bei der Bundestagswahl 2005 zumindest zum Teil auf strategische Wahlentscheidungen zurückzuführen ist, die erst in den letzten Vorwahltagen getroffen wurden und die großen Parteien zugunsten ihrer jeweiligen tatsächlichen oder angestrebten Koalitionspartner Stimmen gekostet haben. Besonders ausgeprägt scheinen solche Wanderungsbewegungen im „schwarz-gelben Lager“ stattgefunden zu haben, dessen Ergebnis die „Prognosen“ auch besonders schlecht vorhersagten. Während diese Interpretationen nur vorläufiger Natur sein können und durch weitere Analysen erhärtet werden müssen, vermitteln die Befunde der in diesem Kapitel dargestellten Analysen auf jeden Fall einen deutlichen Eindruck vom Nutzen tagesgenauer Beobachtung öffentlicher Meinung im Verlauf von Wahlkampagnen mit dem Instrument der RCS-Umfrage.

22 Für die FDP zeigt sich in der *Wahlstreet* ein interessantes Phänomen: Am Wahltag schießt der FDP-Kurs ab ungefähr 16 Uhr auf annähernd 10 Prozent nach oben. Die Vermutung liegt nahe, dass hier ein Händler Profite aus Insiderwissen erzielen wollte, denn gegen 16 Uhr lagen am Wahltag „gut informierten Kreisen“ die ersten Zahlen aus den Wahltagsbefragungen vor.

5 Fazit

Im vorliegenden Beitrag ging es zunächst darum, die zentralen Ideen eines relativ neuen Erhebungsverfahrens zur Erfassung kurzfristiger Prozesse des Wandels der öffentlichen Meinung vorzustellen: der *Rolling Cross-Section*-Umfrage. Im nächsten Schritt wurde diskutiert, wie dieses Design bei der ersten deutschen RCS-Studie umgesetzt wurde, die anlässlich der Bundestagswahl 2005 durchgeführt wurde. Es wurde deutlich, dass die praktische Realisierung einer solchen Umfrage Primärforscher und Feldinstitute in verschiedener Hinsicht vor beträchtliche Herausforderungen stellt und große Ansprüche an die Qualität der Feldarbeit und eine kontinuierliche Studienbetreuung impliziert. Die im dritten Teil des Artikels vorgestellten Befunde aus unserer Studie können jedoch als Beleg dienen, dass sich der Aufwand (und der damit verbundene finanzielle Einsatz) lohnt. An diesen wenigen exemplarischen Befunden, die das analytische Potenzial der Daten bei weitem nicht ausschöpfen, wird unschwer ersichtlich, dass RCS-Daten neue Einblicke in Wahlkämpfe und andere dynamische Phänomene der öffentlichen Meinung ermöglichen.

Vor dem Hintergrund zunehmender Volatilität des politischen Verhaltens sind einfache Querschnittsstudien, auch wenn sie in monatlichen oder gar wöchentlichen Zeitabständen wiederholt werden, aber auch Panel-Studien immer weniger imstande, die Dynamik der Meinungsbildung und Präferenzformierung nachzuzeichnen. Die berichteten Ergebnisse deuten auf erhebliche Fluktuationen in den Wahlabsichten hin, die sich in den letzten Tagen unmittelbar vor der Wahl stark intensivierten. Diese Prozesse dürften zumindest eine Teilantwort auf das Rätsel der schlechten Prognoseleistungen der für die Medien durchgeführten Meinungsumfragen bei dieser Wahl implizieren. RCS-Erhebungen sind besonders geeignet, um Phänomene des „late swing“ aufzudecken, welche noch unmittelbar vor Wahlen zu erheblichen Verschiebungen der Stimmenanteile von Parteien führen können. Lediglich was am Wahltag selbst geschah, entging bei der Bundestagswahl 2005 selbst diesem Instrument. Erste Analysen unserer Nachwahl-Welle deuten darauf hin, dass fast ein Zehntel der Wähler noch nach Ende der RCS-Erhebung eine Wahlentscheidung getroffen oder revidiert hat. Durch die Kombination der Vorwahl-RCS-Erhebung mit einer Nachwahl-Panel-Welle besteht jedoch eine ideale Voraussetzung, um auch diesen Prozessen analytisch auf die Spur zu kommen.

Literatur

- Aarts, K., van der Kolk, H. & Kamp, M. (1999). *Dutch Parliamentary Election Study 1998*. Documentation, Enschede: NIWI/Steinmetz Archive/Dutch Electoral Research Foundation (SKON).
- Allsop, D. & Weisberg, H. F. (1988). Measuring Change in Party Identification in an Election Campaign. *American Journal of Political Science*, 32, 996-1017.
- Asher, H. (1998). *Polling and the Public. What Every Citizen Should Know, 4th edition*. Washington, DC: CQ Press.
- Bartels, L. M. (1988). *Presidential Primaries and the Dynamics of Public Choice*. Princeton/NJ: Princeton University Press.
- Bartels, L. M. (1999). Panel Effects in the American National Election Studies. *Political Analysis*, 8, 1-20.
- Bartels, L. M. (2006). Three Virtues of Panel Data for the Analysis of Campaign Effects. In H. E. Brady & R. Johnston (Eds.), *Capturing Campaign Effects*, forthcoming. Ann Arbor: University of Michigan Press.
- Blais, A., Gidengil, E., Nadeau, R. & Nevitte, N. (2002). *Anatomy of a Liberal Victory: Making Sense of the Vote in the 2000 Canadian Election*. Peterborough: Broadview Press.
- Blakeman, B. W. (1995). Tracking Polls: How to Do Them. *Campaigns & Elections*, 16, 24-25.
- Brady, H. E. (2000). Contributions of Survey Research to Political Science. *PS – Political Science and Politics*, 33, 47-57.
- Brady, H. E. & Johnston, R. (2006). *Capturing Campaign Effects*. Ann Arbor: University of Michigan Press (i. E.).
- Brady, H. E., Johnston, R. & Sides, J. (2006). The Study of Political Campaigns. In H. E. Brady & R. Johnston (Eds.), *Capturing Campaign Effects*, forthcoming. Ann Arbor: University of Michigan Press.
- Brettschneider, F. (1992). Der taktische und rationale Wähler. Über den Einfluß von Wahlumfragen auf das Wählerverhalten bei den Bundestagswahlen 1983 und 1990. *Politische Vierteljahresschrift*, 33, 55-72.
- Brüggelambert, G. (1999). *Entscheidungsverhalten und Institutionen als Informationsträger: Erfahrungen mit Wahlbörsen*. Marburg: Metropolis.
- Campbell, A., Converse, P. E., Miller, W. E. & Stokes, D. E. (1960). *The American Voter*. New York: Wiley.
- Clarke, H. D., Sanders, D., Stewart, M. C. & Whiteley, P. (2004). *Political Choice in Britain*. Oxford: Oxford University Press.

- Clarke, H. D., Sanders, D., Whiteley, P. & Stewart, M. (2006). Campaign Dynamics and Campaign Effects in the 2001 British General Election. In A. McCutcheon (Eds.), *The Science of Pre-Election Polling*, forthcoming. Lincoln: University of Nebraska Press.
- Cleveland, W. S. (1994). *The Elements of Graphing Data, revised edition*. Murray Hill/NJ: At&T Bell Labs.
- Clymer, A. (2006). *NAES 04. National Annenberg Election Study*. Retrieved April 7, 2006, from <http://www.annenbergpublicpolicycenter.org/naes/>
- Das Parlament (2005). *Der „Umfragesieger-Besieger“*. Retrieved April 7, 2006, from <http://www.das-parlament.de/2005/36/Thema/003.html>
- Dichter Nebel (2005). *Spiegel*, 55.
- Dobrzynska, A., Blais, A. & Nadeau, R. (2003). Do the Media Have a Direct Impact on the Vote? The Case of the 1997 Canadian Election. *International Journal of Public Opinion Research*, 15, 27-43.
- Dörner, C. (1998). Do Media Matter? – Zur Reichweite des Agenda Settings. In C. Dörner & Klaudia Erhardt (Eds.), *Politische Meinungsbildung und Wahlverhalten. Analysen zum „Superwahljahr“ 1994* (pp. 119-149). Wiesbaden: Westdeutscher Verlag.
- Elff, M. (2006). Nichtparametrische und semiparametrische Regressionsverfahren. In J. Behnke, T. Gschwend, D. Schindler & K. Schnapp (Eds.), *Methoden der Politikwissenschaft*, forthcoming. Baden-Baden: Nomos.
- Erhardt, K. (1998). Die unentschlossenen Wähler als Motor zyklischer Wahlabsichtverläufe. Eine Untersuchung am Beispiel der Rückgewinnung der Wähler durch die CDU im Frühjahr 1994. In C. Dörner & K. Erhardt (Eds.), *Politische Meinungsbildung und Wahlverhalten. Analysen zum "Superwahljahr" 1994* (pp. 15-118). Wiesbaden: Westdeutscher Verlag.
- Farrell, D. M. & Schmitt-Beck, R. (2002). *Do Political Campaigns Matter? Campaign Effects in Elections and Referendums*. London/New York: Routledge.
- Fiorina, M. P. (1981). *Retrospective Voting in American National Elections*. New Haven: Yale University Press.
- Fournier, P., Nadeau, R., Blais, A., Gidengil, E. & Nevitte, N. (2004). Time-of-voting decision and susceptibility to campaign effects. *Electoral Studies*, 23, 661-681.
- Fox, J. (2000). *Nonparametric Simple Regression: Smoothing Scatterplots*. Thousand Oaks: Sage.
- FTD (2005). *Gallup-Institut kritisiert deutsche Meinungsforscher*. Retrieved April 7, 2006, from <http://www.ftd.de/pw/de/23035.html>
- Gabler, S. & Häder, S. (1997). Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. *ZUMA-Nachrichten*, 41, 7-18.

- Gidengil, E. & Dobrzynska, A. (2003). *Using a rolling cross-section design to model media effects: The case of leader evaluations in the 1997 Canadian elections*. Paper prepared for delivery at the Annual Meeting of the APSA, Philadelphia.
- Glagow, H. & Bühler, E. (1987). *C.A.T.I. – Das Antwortverhalten*. Sample Workshop Nr. 2.
- Güllner, M. (2000). Methodische Entwicklungen in der Empirischen Wahlforschung. In M. Klein, W. Jagodzinski, E. Mochmann & D. Ohr (Eds.), *50 Jahre Empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten* (pp. 564-583). Wiesbaden: Westdeutscher Verlag.
- Güllner, M., Dülmer, H., Klein, M., Ohr, D., Quandt, M., Rosar, U. & Klingemann, H. (2005). *Die Bundestagswahl 2002. Eine Untersuchung in Zeiten hoher politischer Dynamik*. Wiesbaden: VS-Verlag.
- Häder, S. & Gabler, S. (1998). Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In S. Häder, S. Gabler & J. H.P. Hoffmeyer-Zlotnik (Eds.), *Telefonstichproben in Deutschland* (pp. 69-88). Opladen: Westdeutscher Verlag.
- Hillygus, D. S. & Jackman, S. (2003). Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy. *American Journal of Political Science*, 47, 583-596.
- Hofmann-Göttig, J. (2005). *Umfragen zur Bundestagswahl vom 18. September 2005: Was taugte die „Sonntagsfrage“?* Retrieved April 7, 2006, from http://www.mbfj.rlp.de/Wir_ueber_uns/downloads/Umfragen_zur_Bundestagswahl_2005.pdf
- Holbrook, T. M. (1996). *Do Campaigns Matter?* Thousand Oaks: Sage.
- Huber, J. (2002). *Wahlbörsen: Preisbildung auf politischen Märkten zur Vorhersage von Wahlergebnissen*. Hamburg: Kovac.
- Johnston, R. (2001). Capturing Campaigns in National Election Studies. In E. Katz & Y. Warshel (Eds.), *Election Studies. What's Their Use?* (pp. 149-172). Boulder/Col.: Westview.
- Johnston, R., Blais, A., Brady, H. E. & Crête, J. (1992). *Letting the People Decide. Dynamics of a Canadian Election*. Stanford/CA: Stanford University Press.
- Johnston, R., Blais, A., Gidengil, E. & Nevitte, N. (1996). *The Challenge of Direct Democracy. The 1992 Canadian Referendum*. Montreal: McGill-Queen's University Press.
- Johnston, R. & Brady, H. E. (2002). The rolling cross-section design. *Electoral Studies*, 21, 283-295.
- Johnston, R., Hagen, M. G. & Jamieson, K. H. (2004). *The 2000 Presidential Election and the Foundations of Party Politics*. Cambridge/Mass: Cambridge University Press.
- Lazarsfeld, P. F., Berelson, B. & Gaudet, H. (1944). *The People's Choice. How the Voter Makes Up his Mind in a Presidential Campaign*. New York/London: Columbia University Press.
- Mosmann, H. (1998). *Das Antwortverhalten PAPI – CAPI – CATI*. Intra Workshop Nr. 6.

- Nevitte, N. & Blais, A., Gidengil, E. & Nadeau, R. (2000). *Unsteady State: The 1997 Canadian Federal Election*. Don Mills: Oxford University Press.
- Northrup, D. (1989). *The 1988 National Election Study. Technical Documentation*. Toronto: Institut for Social Research, York University. Retrieved April 7, 2006, from http://www.isr.yorku.ca/download/CES/CES88/Techdoc_88.pdf
- Northrup, D. (1994). *The 1993 Canadian Election Study. Technical Documentation*. Toronto: Institut for Social Research, York University. Retrieved April 7, 2006, from <http://www.isr.yorku.ca/download/CES/CES93/TECHDOC.PDF>
- Northrup, D. (1998). *The 1997 Canadian Election Survey. Technical Documentation*. Toronto: Institut for Social Research, York University. Retrieved April 7, 2006, from http://www.isr.yorku.ca/download/CES/CES97/TECH_ENG.PDF
- Northrup, D. (2003). *The 2000 Canadian Election Survey. Technical Documentation*. Toronto: Institut for Social Research, York University. Retrieved April 7, 2006, from http://www.isr.yorku.ca/download/CES/CES00/CES_TD_001.pdf
- NZZ (2005). *Deutschlands Wahlforscher als Verlierer*. Retrieved April 7, 2006, from http://www.nzz.ch/dossiers/Wahlen_Deutschland/resultate/2005/09/21/al/articleD5S9A.html
- Romer, D., Kenski, K., Waldman, P., Adasiewicz, C. & Jamieson, K. H. (2004). *Capturing Campaign Dynamics: The National Annenberg Election Survey. Design, Method, and Data*. New York/Oxford: Oxford University Press.
- Sanders, D. J., Whiteley, P. F., Clarke, H. D., Stewart, M. & Winters, K. *The British Election Study. 2005 Rolling Campaign Survey Data*. Retrieved April 7, 2006, from <http://www.essex.ac.uk/bes/2005/ircp.htm>
- Schmitt-Beck, R. (1996). Mass Media, the Electorate, and the Bandwagon. A Study of Communication Effects on Vote Choice in Germany. *International Journal of Public Opinion Research*, 8, 266-291.
- Schmitt-Beck, R. (2005). *Kampagnen-Dynamik. Eine Rolling Cross-Section/Panel-Studie zu den Wirkungen des Wahlkampfes bei der vorgezogenen Bundestagswahl 2005*. Projektantrag an die Deutsche Forschungsgemeinschaft, Institut für Politikwissenschaft der Universität Duisburg-Essen.
- Schmitt-Beck, R. & Weick, S. (2001). Die dauerhafte Parteiidentifikation – nur noch ein Mythos? Eine Längsschnittanalyse zur Identifikation mit politischen Parteien in West- und Ostdeutschland. *Informationsdienst Soziale Indikatoren*, 26, 1-5.
- Schmitt-Beck, R., Weick, S. & Christoph, B. (2006). Shaky Attachments: Individual-Level Stability and Change of Partisanship among West German Voters, 1984-2001. *European Journal of Political Research*, 35 (i. E.).
- Schnell, R. (1994). *Graphisch gestützte Datenanalyse*. München: Oldenbourg.
- Schoen, H. (2005). *Wahlkampfforschung*. In J. W. Falter & H. Schoen (Eds.), *Handbuch Wahlforschung* (pp. 503-542). Wiesbaden: VS-Verlag.

- Shaw, D. R. (2000). Campaign Events, the Media and the Prospects of Victory: The 1992 and 1996 US Presidential Elections. *British Journal of Political Science*, 30, 259-289.
- Stöss, R. (1997). *Stabilität im Umbruch. Wahlbeständigkeit und Parteienwettbewerb im "Superwahljahr" 1994*. Opladen: Westdeutscher Verlag.
- Vowles, J., Aimer, P., Banducci, S. & Karp, J. (1998). *Voters' Victory? New Zealand's First Election Under Proportional Representation*. Auckland: Auckland University Press.
- Vowles, J., Aimer, P., Banducci, S., Karp, J. & Miller, R. (2004). *Voters' Veto. The 2002 Election in New Zealand and the Consolidation of Minority Government*. Auckland: Auckland University Press.
- Vowles, J., Aimer, P., Karp, J., Banducci, S., Miller, R. & Sullivan, A. (2002). *Proportional Representation on Trial*. Auckland: Auckland University Press.
- Whiteley, P. F. (1988). The Causal Relationships between Issues, Candidate Evaluations, Party Identification, and Vote Choice – the View from "Rolling Thunder". *Journal of Politics*, 50, 961-984.
- Wüst, A. (2003). Stimmung, Projektion, Prognose? In A. Wüst (Ed.), *Politbarometer* (pp. 83-107). Opladen: Leske + Budrich.
- ZDF (2006). *Politbarometer auf einen Blick*. Retrieved April 7, 2006, from <http://www.zdf.de/ZDFde/inhalt/12/0,1872,2337900,00.html>

Korrespondenzadressen

Prof. Dr. Rüdiger Schmitt-Beck
Institut für Politikwissenschaft
Universität Duisburg-Essen – Campus Duisburg
Lotharstraße 65
47057 Duisburg
E-mail: ruediger.schmitt-beck@uni-due.de

Dr. Christian Holst
Director Public Affairs /
Politik & Sozialforschung
IPSOS GmbH
Papenkamp 2-6
23879 Moelln
E-mail: Christian.Holst@Ipsos.de

OVERREPORTING: EIN VERNACHLÄSSIGTES PROBLEM UND DIE SCHWIERIGKEIT VON GEGENMASSNAHMEN

OVERREPORTING: A NEGLECTED ISSUE IN GERMAN- LANGUAGE LITERATURE AND THE PROBLEM OF COUNTERMEASURES

SIBYLLE HARDMEIER & MARIE-CHRISTINE FONTANA

Overreporting wird in der deutschsprachigen Methodenliteratur kaum behandelt – zu Unrecht, wie ein Blick auf die empirischen Daten zeigt. Der vorliegende Artikel hat sich daher zum Ziel gesetzt, dieses in Vergessenheit geratene Problem in Erinnerung zu rufen, und geht auf drei Ebenen vor: Im ersten Teil präsentieren wir einen Überblick über den beachtlichen Literaturkorpus im englischsprachigen Raum und nehmen eine Systematisierung entlang methodischer Problemstellungen und theoretischer Begründungen des Phänomens vor – der sozialen Erwünschtheit einerseits und dem Misremembering andererseits, das als Erinnerungsproblem oder als „source confusing“ verstanden werden kann. Im zweiten Teil widmen wir uns den zwei theoretischen Erklärungssets anhand empirischer Daten aus der Schweiz. Mit der Hilfe von Aggregatdatenanalysen sowie dreier Feldexperimente nehmen wir erste explorative Hypothesentests vor und analysieren den Effekt theoretisch hergeleiteter Maßnahmen auf der Ebene des Fragebogens (Wording). Es zeigt sich, dass Gegenmaßnahmen zur sozialen Erwünschtheit mit kleinen Modifikationen im Wording schwierig zu erreichen sind und sogar kontraproduktiv sein können. Ein Effekt des Treatments zum Misremembering ist bei großen Zeitabständen zwar vorhanden, jedoch bleibt die kausale Begründung offen und schwierig. Daher schlagen wir im letzten Teil weitere Maßnahmen für künftige Untersuchungen vor: Ein Bogus-pipeline-Experiment, Kontrollen über Erwünschtheits- bzw. Erinnerungsskalen, eine Meta-Analyse und Validitätsstudien für jene Länder, die noch keine haben.

Overreporting is hardly discussed in German-language methodology literature – hardly justifiable, as a glance at the empirical data shows. The intention of the present article is therefore to bring this forgotten problem back to memory. We proceed at three levels: the first part presents an overview of the considerable body of literature

found in English-speaking countries and classifies it according to methodological approaches and theoretical explanations of the phenomenon – social desirability on the one hand and misremembering on the other, which can be interpreted as a memory failure or as “source confusing”. The second part is devoted to two theoretical sets of explanations, using empirical data from Switzerland. We conduct some initial explorative hypotheses tests based on aggregate data analyses and three field experiments, and examine the effect of theoretically deduced measures at the level of questionnaires (wording). This reveals that it is difficult to create countermeasures to social desirability with minor modifications of the wording and that it may even be counter-productive. Despite treatment of misremembering having an effect when time spans are sufficiently large, the causal explanation remains unclear and difficult. The last part therefore proposes additional measures for future investigations: a bogus pipeline experiment, checks on desirability and memory scales, a meta-analysis and validity studies for countries still lacking them.

1 Einleitung

In einer Zeit, in der wir beinahe täglich mit den Ergebnissen von Befragungsdaten konfrontiert werden, geht leicht vergessen, dass die Demoskopie ein sensibles Erhebungsverfahren darstellt und mit Selektions- ebenso wie mit Messproblemen (Groves 1989) konfrontiert ist. Am Beispiel des so genannten Overreporting – der Tatsache also, dass bei Wahlbefragungen der Anteil jener, die angeben, an der Wahl teilgenommen zu haben, höher ist als die tatsächliche Wahlbeteiligung – ist das besonders dokumentiert. Die höhere, in den Befragungsdaten angezeigte Partizipationsrate wird nämlich, abgesehen von Referenzproblemen in der offiziellen Statistik¹ und allfälligen Stimuluseffekten in Panels, im Wesentlichen auf zwei unterschiedliche, allenfalls jedoch gleichzeitig auftretende Ursachen zurückgeführt:

- dass Nichtwählende in der gezogenen Stichprobe untervertreten sind (Sampling-Fehler);
- und dass einige Befragte eine Fehlantwort zu ihrem vermeintlichen Teilnahmeverhalten geben, womit das Overreporting zu einem speziellen Fall von Misreporting wird (Messfehler).

1 Die offizielle Statistik kann zu viele Wahlberechtigte ausweisen und damit den Nichtwähleranteil überschätzen, wenn sie Wegzüge, Todesfälle oder Ähnliches nicht auszuweisen vermag. Vgl. Clausen (1968); Kleinhenz (1995); Lavies (1973); Presser et al. (1990).

Bereits bei der Durchführung der ersten repräsentativen Befragungen in den USA wurde die politische Demoskopie mit diesem ernsthaften Validitätsproblem² konfrontiert. Wie eine Recherche in JSTOR zeigt,³ beschäftigt sich die amerikanische Zunft auch heute noch mit diesem Problem. Im Gegensatz dazu scheint die Auseinandersetzung – obwohl auch im Kontext europäischer Wahlen früh konstatiert⁴ – im deutschsprachigen Raum deutlich weniger intensiv. Eine entlang identischer Suchkriterien⁵ wie bei JSTOR durchgeführte Literaturrecherche in den einschlägigen wissenschaftlichen deutschsprachigen Fachzeitschriften⁶ generiert keinen einzigen Hit.⁷

Ist dieses Schweigen gerechtfertigt und sind die Mahnrufe von frühen Protagonisten der deutschen (Kaase 1973) oder schweizerischen Wahlforschung (Steiner 1965) heute obsolet? Oder müssen wir vielmehr eine wichtige Forschungslücke benennen? Ein flüchtiger Blick in die schweizerischen und deutschen Politdaten gibt eine erste Antwort auf diese Fragen: Im Politbarometer 2002 gaben 89.8 Prozent der Befragten im September 2002 an, sie würden sicher zur Wahl gehen oder hätten bereits brieflich gestimmt. Weitere 6.4 Prozent hielten fest, sie würden wahrscheinlich teilnehmen. Gemessen an der faktischen Beteiligung vom 22. September 2002 von 79.1 Prozent ist damit von einem Overreporting von 10.7 Prozentpunkten auszugehen.⁸ Gleichtags wurden die Schweizer Stimmberechtigten zu einer Referendumsabstimmung an die Urnen gerufen. In der VOX-Nachbefragung gaben 61.2 Prozent der Befragten an, sie hätten ihre Stimme abgegeben; die effektive Beteiligung lag jedoch nur bei 45.2 Prozent, womit sich ebenfalls ein deutliches Overreporting abzeichnete (15 Prozentpunkte).

2 In den meisten Overreporting-Studien wird aufgrund der Annahme von systematischen Messfehlern das Problem der Validität in der Tat als bedeutender angesehen als jenes der Reliabilität (Carmines & Zeller 1979: 13 ff). Kritisch dazu äußern sich Andersson & Granberg (1997: 127).

3 Eine Recherche mit dem Suchbegriff „Overreporting“ in den einschlägigen Journals der Sozialwissenschaft hat nicht weniger als 64 Zeitschriftenartikel zu Tage gefördert; wobei 14 davon als eigentliche Overreporting-Studien zu taxieren sind.

4 Swaddle & Heath 1989; Kleinhenz 1995; Rattinger & Wagner 2003; van der Kolk & Aarts 2001, Granberg & Holmberg 1991 sowie Riklin & Kley 1981.

5 Also: <overr> sowie <antw>, <antwortstil>, <messfehl>, <erinnerung>, <response>.

6 Kölner Zeitschrift für Soziologie und Sozialpsychologie (Gesamtinhaltsverzeichnis der Titel 2003-1996); Politische Vierteljahresschrift (Jahresinhaltsverzeichnis der Titel 2003-2000); ZA-Informationen (Integrale Ausgaben ab Nr. 37 sowie Titel und Abstracts der Nummern 1-36); ZUMA-Nachrichten (Gesamtverzeichnis der Titel 2003-1977) sowie ZUMA-Arbeitsberichte (Gesamtverzeichnis der Titel 2003-1980).

7 Dabei kann das nicht mit einer Abneigung gegenüber anglistischen Termini begründet werden: Zum einen finden andere Anglizismen wie „Survey“, „Bias“ oder „Response“ durchaus Anklang. Zum anderen wird der Fachterminus „Overreporting“ in Buchpublikationen durchaus verwendet (Kaase & Bauer-Kaase 1998: 87). Allerdings gilt es festzuhalten, dass eines der wichtigsten Survey-Methodenbücher aus Deutschland (Schnell 1997) sich ‚nur‘ mit dem Unit Response beschäftigt und den Ausfall von Nichtwählern nicht speziell untersucht.

8 Politbarometer West 2002 ZA-Nr. 3849, S. 25 sowie Statistisches Bundesamt, Bundestagswahl 2002, Heft 3, S. 6.

Der vorliegende Artikel widmet sich daher dem Phänomen des Overreporting – nachfolgend mit OR abgekürzt – und verfolgt folgende Ziele: Wir möchten erstens den aktuellen Forschungsstand aufarbeiten und dem interessierten Publikum zugänglich machen. Die beachtliche Anzahl Publikationen aus dem englischsprachigen Raum werden in Kapitel 2 anhand theoretischer und methodischer Kriterien systematisch diskutiert. Zudem werden die Resultate einiger empirischer Primär- und Sekundärdatenerhebungen präsentiert und im aktuellen Forschungsstand verortet, wobei der Effekt des Wording im Vordergrund steht. Daten und Experimente aus der Schweiz, die in Kapitel 3 präsentiert werden, bestätigen den Eindruck, dass Gegenmaßnahmen in diesem Bereich schwierig und zwiespältig sind. Vor diesem Hintergrund ziehen wir in Kapitel 4 einige Folgerungen und präsentieren Anregungen für weitergehende Untersuchungen.

2 Forschungsstand

In den USA ist die wissenschaftliche Auseinandersetzung mit dem OR so alt wie die Wahlforschung. Nachdem bei den Wahlgängen von 1936 und 1948 die Wahlprognosen kräftig daneben lagen, begann die Disziplin, Validitätsprobleme von Wahlbefragungen theoretisch zu erörtern, das Phänomen des OR deskriptiv zu erfassen und empirische Validitätsstudien durchzuführen. Diese dienten dazu, Profile der Misreporter zu erstellen und kausale Erklärungen zu präsentieren. Gleichzeitig begann man mögliche Korrekturmechanismen zu entwickeln und die allfällige Verzerrung von Modellschätzungen zu erörtern.

2.1 Overreporting als Misreporting

Bereits die ersten deskriptiven Untersuchungen zum Phänomen des OR konnten einen recht konsistenten Befund festhalten: Sobald es nicht um statische Faktbeschreibungen wie den Besitz eines Telefons, sondern um „ex-post“ Angaben zu einem Verhalten geht, das normativ aufgeladen ist, sind die Befragungsdaten in eine Richtung verzerrt. Auch bei den amerikanischen Wahlstudien wurde dieses Phänomen von Beginn an beobachtet und die Abweichungen auf aggregiertem Niveau bewegen sich zwischen 13 und 15 Prozentpunkten (Presser et al. 1990). Allerdings ist aufgrund des Vergleichs ‚Befragungsdaten – offizielle Statistik‘ das Phänomen des Overreporting unterbeleuchtet. Daher wurden in den USA schon früh so genannte Validitätsstudien realisiert, in welchen die selbst bekundete Partizipation mit offiziellen Daten aus den Wahlbüros validiert wurde. Der Vorteil dieses Zugangs liegt auf der Hand: Erst diese Analysen konnten bestätigen, dass das auf aggregiertem Niveau beobachtete Phänomen des OR in vielen Fällen ein Messproblem darstellt und mit einem Misreporting auf individueller Ebene, also bei den Befragten, einhergeht.

Dabei machten erste und thematisch breit angelegte Studien vorerst auf die Verbreitung des Phänomens bei so genannten self-reports in Befragungen aufmerksam. Misreporting findet bei verschiedenen Themen (Cahalan 1968; Parry & Crossley 1950) wie auch unterschiedlichen Partizipationsformen (Volgy & Schwarz 1984) statt und wird offensichtlich durch eine gemeinsame Klammer zusammengehalten: Es tritt auf, sobald die Themen normativ aufgeladen und die entsprechenden Fragen als sensitiv wahrgenommen werden (Renzetti & Lee 1993). Daher kann auch eine Grundregel festgehalten werden (Tourangeau et al. 2000: 269 ff.; Belli et al. 2001; Sigelman 1982): Befragte neigen zu einem Underreporting bei sozial unerwünschten und zu einem Overreporting bei sozial erwünschten Themen. Entsprechend sind Fragen zur politischen Partizipation nicht nur besonders anfällig für diese Verzerrung (Sudman & Bradburn 1974: 59), sondern die Richtung des Misreporting entpuppt sich – von einigen wenigen Ausnahmen abgesehen⁹ – als „Einbahnstraße“ (Eilfort 1994: 90): Das Misreporting ist ein Overreporting.

Allerdings blieben trotz einer frühen Mahnung von Anderson und Silver (1986: 775) die Berechnungsweisen des OR in diesen Validitätsstudien sehr heterogen und zuweilen wenig überzeugend. Denn oft wird für die Berechnung der Anteile nicht der richtige Nenner gewählt und die eigentliche „population at risk“ gerät aus dem Blickfeld:¹⁰ jene Nichtwählenden nämlich, welche unter Umständen sozial stigmatisiert werden, wenn sie nicht teilnehmen. Wenn man also der Empfehlung von Anderson und Silver folgt und misst, wie viele von den validierten Nichtwählenden behaupten, gewählt zu haben, dann ist das OR beträchtlich. So liegt der Wert der wohl umfassendsten Validitätsanalyse für die ANES zwischen 1964 und 1988 nicht wie berichtet bei rund 14 Prozent (Traugott 1989: 8), sondern bei rund 25 Prozent. Bei Swaddle & Heath käme man entsprechend auf 24.9 Prozent und würde sich wiederum auf ähnlich hohem Niveau wie bei Sigelman (1982: 49), Andersson & Granberg (1997), Granberg & Holmberg (1991) oder Karp & Brockington (2005 forthcoming) bewegen. Das Misreporting scheint also nicht nur ein pankulturelles Phänomen darzustellen, sondern bewegt sich auch auf hohem Niveau.

9 Auf wichtige Ausnahmen und damit systematisches Underreporting bei der Teilnahmefrage – wenn „low-salience“ Wahlen zeitgleich zu „high-salience“ Wahlen stattfinden – haben vor allem Adamanly & Shelley (1980) aufmerksam gemacht. Jennings hat in seiner nach Staaten differenzierten Studie festgestellt, dass in Alaska konstant ein Underreporting stattfindet (siehe Presser et al. 1990: 3).

10 Einige Autoren definieren nämlich Misreporting als den Gesamtanteil unkorrekter Antworten (Steiner 1965; Volgy & Schwarz 1984; Cahalan 1968). Andere Autorinnen wie zum Beispiel Sigelman (1982) oder Traugott & Katosh (1979) referieren den Anteil unter jenen, die eine Teilnahme behauptet haben, tatsächlich aber der Urne fernblieben.

2.2 Das Profil der Overreporter

Wenn Validitätsstudien neben der Teilnahmefrage zusätzliche soziodemographische Variablen erfassen, dann können sie einen Beitrag zur Profilbildung der Overreporter liefern. Ein Blick auf die Literatur zeigt allerdings, dass die Erstellung eines solchen Sozialprofils trotz einer Vielzahl realisierter Studien nur beschränkt möglich ist. Neben den Zugangsweisen¹¹ sind auch die Befunde disparat. Während beispielsweise die einen Schwarze oder Nichtweiße eher als Overreporter identifizieren,¹² wird dies von Anderson & Silver (1986) methodisch in Frage gestellt und von Bernstein et al. (2001) differenziert. Da zudem überraschend oft die jüngeren Stimmberechtigten – von denen man vermuten könnte, sie hätten die soziale Norm der Teilnahme weniger verinnerlicht – zu den Overreportern zählen,¹³ trifft die Einschätzung, dass die Overreporter eher den Wählenden als den Nichtwählenden ähneln, nur beschränkt zu. In der Tendenz – und vor dem Hintergrund von methodisch sophistizierteren Analysen (Anderson & Silver 1986; Silver et al. 1986; Bernstein et al. 2001) – lässt sich aber die These aufstellen, dass es vor allem die politisch und gesellschaftlich Involvierten sind, welche zum OR neigen.

2.3 Theoretische Begründungen: Soziale Erwünschtheit und Misremembering

Die regelmäßig durchgeführten Validitätsstudien der ANES haben den Eindruck erhärtet, dass es sich beim OR der Wahlteilnahme um ein Messproblem handelt. Als wichtigste Erklärung dafür wurde die soziale Erwünschtheit aufgeführt. Dieser formale Antwortstil (Crowne & Marlowe 1964) beschreibt „the tendency of people to deny socially undesirable traits and to ‘admit’ socially desirable ones” (Phillips 1973: 20). Das heißt mit anderen Worten, der Begriff des “Lügens” vermag zwar das Antwortverhalten der Befragten faktisch zu umschreiben, hinsichtlich der dahinter liegenden Motivation ist er aber wenig erhellend (Mummendey 1999: 160). Die spezifische Strategie der Selbstpräsentation in Form von OR hat vielmehr mit sozialen Normen und dem Bedürfnis nach sozialer Anerkennung zu tun (Scholl 1993: 67 ff.); wobei diese Strategie der „intentional deception“ (Belli et al. 2001: 494) manchmal eher als Persönlichkeitszug, dann wiederum eher als Situationsreaktion verstanden werden kann.¹⁴

11 Dabei ist vor allem auf zwei Differenzen hinzuweisen: Weil erstens, wie in Abschnitt 2.1 ausgeführt, unterschiedliche Populationen bzw. Nenner betrachtet werden, basieren die Sozialprofile auf unterschiedlichen Gruppenvergleichen. Zweitens ist das analytische Vorgehen unterschiedlich. Während sich die einen bei der Profilerstellung auf deskriptive Zugänge beschränken, versuchen die anderen, das Profil der Overreporter kausal zu erklären (Traugott & Katosh 1979; Katosh & Traugott 1981).

12 Abramson & Claggett (1986); Belli et al. (2001); Hill & Hurley (1984); Sigelman (1982).

13 Bernstein et al. 2001; Cahalan 1968; Traugott & Katosh 1979; Wolfinger & Rosenstone 1980.

14 Eine entsprechende Unterteilung der Motive findet sich auch in den stärker von der Rational-choice-Theorie inspirierten Ansätzen der Wert-Erwartungstheorie (Esser 1991; Tourangeau et al. 2000).

Allerdings entpuppt sich vor diesem Hintergrund das response set „Overreporting“ auch als ein recht volatiles und kontextabhängiges Konstrukt, das Prognosen darüber, wer zu den Overreportern gehört, nicht einfach macht. Insgesamt ist sich aber die Literatur – auch vor dem Hintergrund der Ausführungen unter Punkt 2.3 – einig, dass es eher die Personen mit höherer Bildung und stärkerem Involvierungsgrad sind. Diese Personengruppen seien sich der gesellschaftlich ‚korrekten‘ Antworten stärker bewusst (Silver et al. 1986: 623). Die Beteiligung an Wahlen und Abstimmungen ist eben nicht nur ein freier Entscheid, sondern stellt auch eine sozial vermittelte Bürgerpflicht dar. OR aus Gründen der sozialen Erwünschtheit setzt also ein entsprechendes Pflichtbewusstsein voraus. Stark verknüpft, wenn auch nicht immer expliziert, mit der sozialen Erwünschtheitstheorie sind Vermutungen über den Einfluss der Erhebungsmethode auf das OR sowie über den Wandel des OR im Laufe der Zeit:

Allgemein wird angenommen, dass die Erhebungsmethode mit dem höheren Grad an Anonymität und weniger direktem Kontakt mit Interviewern – das Telefoninterview – weniger soziale Erwünschtheit allgemein und damit auch weniger OR provoziert. Während Untersuchungen zu sensitiven Fragen diese These bestätigen (de Leeuw & van der Zouwen 1988; Tourangeau et al. 2000: 295) und weitgehend erhärtet ist, dass anonyme Verfahren die Verzerrung durch soziale Erwünschtheit reduzieren (Richman et al. 1999: 756), ist der Literaturkorpus für das OR noch schmal und liefert disparate Ergebnisse.¹⁵ Einige Studien können zumindest einen schwachen, positiven Effekt der telefonischen Erhebungsmethode nachweisen (Locander et al. 1976; Rogers 1976). Andere können das nicht bestätigen (Abramson & Claggett 1986: 416) bzw. sehen keine Effekte der Erhebungsmethode (Voogt & Saris 2003: 170 f.).¹⁶

Für moderne Demokratien wird häufig ein Prozess der Individualisierung diagnostiziert, der sich in der Wahlforschung in abnehmenden Parteibindungen und Teilnahmeraten niederschlägt. Die Beteiligung an Wahlen und Abstimmungen – so die allgemeine Diagnose – wird stärker als freiwillige Handlung und weniger als soziale Pflicht angesehen. Folglich sollte auch die soziale Erwünschtheit der politischen Teilnahme abgeschwächt worden sein.¹⁷ Allerdings bestätigen die bisher publizierten Daten eher das Gegenteil.

15 Das Argument geht auf Sudman & Bradburn (1974: 9 f.) zurück. Der National Household Survey on Drug Abuse zum Beispiel verwendet deshalb selbst administrierte, schriftliche Befragungen. Zu Ausnahmen beim OR vgl. auch Voogt & Saris (2003: 170). Es stellt sich vor allem die Frage, ob schriftliche oder Online-Erhebungen neben der ‚Anonymität‘ nicht auch einen höheren Grad an Verbindlichkeit vermitteln.

16 Ein Blick auf das Design der Studie (ebd: 168) lässt allerdings vermuten, dass hier Sampling- und Erhebungseffekte zusammenwirken.

17 In einer Querschnittsanalyse bestätigen Hill & Hurley (1984) diese These indirekt. Aufrichtige Nichtwähler haben einen schwächer ausgeprägten Sinn für die Staatsbürgerpflicht.

Während Belli et al. (2001: 494) tatsächlich eine Abnahme im OR konstatieren, gibt es überzeugende, wenn auch nicht ungeteilte Hinweise dafür, dass diese Diskrepanz zwischen offizieller Statistik und Selbstdeklaration in der jüngsten Zeit eher zugenommen hat (Bernstein et al. 2001; Burden 2003; Burden 2000; van der Kolk & Aarts 2001).¹⁸ Eine nahezu konstante Diskrepanz zwischen faktischer und bekundeter Teilnahme geht nach Burden mit einer abnehmenden Wahlbeteiligung und somit einem steigenden OR einher – unabhängig davon, wie man die Wahlbeteiligung misst.

Die kognitive Wende in der Sozialwissenschaft und die Sensibilisierung dafür, dass Prozesse der Informationsverarbeitung ebenso wie die Beantwortung von Survey-Fragen komplexe und mehrstufige Verfahren darstellen, haben in der Befragungsforschung das Interesse für so genannte Erinnerungs- (recall) und Abrufungsprobleme (retrieval) geweckt (Tourangeau et al. 2000: 7 ff.). Das hat auch Einfluss auf die OR-Debatte. Ergänzend zur Intentional-deception-Hypothese wird in jüngster Zeit verstärkt die Misremembering-Hypothese ins Feld geführt (Belli et al. 2001: 494). Mit dem differenzierteren Wissen über Gedächtnisstrukturen und der Sensibilisierung dafür, dass der Prozess des Retrievals durch verschiedene Faktoren gestört werden kann, erscheint das Antworten auf Fragen keineswegs als banale Angelegenheit. Deshalb begann man neben der Frage zum Stimmverhalten auch die Partizipationsfrage vor diesem kognitionspsychologischen Hintergrund zu untersuchen (Abelson et al. 1992; Belli et al. 1999 bzw. 2001).

Dabei können ‚simple‘ Erinnerungsprobleme ausschlaggebend sein – insbesondere dann, wenn der Erinnerungsgegenstand weit zurückliegt. Allerdings ist es zumeist nicht die Zeitspanne allein, welche einen negativen Einfluss hat, sondern die Tatsache, dass in der Zwischenzeit andere oder sogar ähnliche Ereignisse stattgefunden haben. Nicht Erinnerung, sondern Retrievement wird damit die kritische Größe, weil das so genannte „source monitoring“ gestört wird (Belli et al. 2001: 480; Tourangeau et al. 2000: 83). Das bedeutet indessen auch, dass sich das Profil der potenziell Anfälligen ändern sollte: Während einfache Erinnerungsprobleme bei den weniger Gebildeten und politisch weniger Interessierten größer sein sollten (Adamany & Shelley 1980), sind vor allem die Involvierten vom source confusing betroffen. Nur jene, welche die Absicht hatten an der Wahl teilzunehmen, können Intentionen und Handlungen unter Umständen verwechseln; und nur wenn andere politische Partizipationsformen in der Zwischenzeit stattgefunden haben, kann Konfusion entstehen.

18 Burden kann die Kritik (Martinez 2003; McDonald 2003) an seinem Artikel aus dem Jahre 2000, wonach die vermeintliche Zunahme lediglich auf die unterschiedlichen Grundgesamtheiten der Wahlstudien und offiziellen Statistiken zurückzuführen sei, empirisch widerlegen.

Aufgrund dieser Tatsache und im Wissen darum, dass das Misreporting nicht zufällig, sondern gerichtet ist, schlagen Belli und seine Kolleginnen (Belli et al. 2001; Belli et al. 1999) eine Kombination der beiden Erklärungsansätze vor. Die soziale Erwünschtheit spiele vor allem dann eine Rolle, so ihre Annahme, wenn Erinnerungsprobleme auftreten. Andererseits müssten Erinnerungsprobleme allein, ohne Kombination mit sozialer Erwünschtheit, genau so oft zu Under- wie zu Overreporting führen. Da dies bekanntermaßen nicht der Fall ist, sei ein Zusammenspiel der beiden Erklärungsfaktoren plausibel (Belli et al. 1999: 91).

2.4 Einfluss auf Modellschätzungen und Gegenmaßnahmen

Nachdem sich die ANES zunächst prioritär den Profilen und Ursachen des OR gewidmet hatte, provozierte nicht zuletzt das Streichen der Validitätsanalysen im Jahre 1990 einen Wandel in der Forschungsagenda. So erlaubten es die mittlerweile großen Zeitreihen, den oben besprochenen Einfluss von Kontexteffekten, von neuen Erhebungsmethoden oder abnehmenden Parteibindungen und Beteiligungsraten zu erörtern. Darüber hinaus wurde der Frage nach möglichen Effekten auf Modellschätzungen wieder stärkere Aufmerksamkeit zuteil und die Befunde verweisen vermehrt in die Richtung, die bereits Hyman (1945: 559) aufgezeigt hatte: Die Parameter der Wahl- und Abstimmungsforschung sowie Modellschätzer von multivariaten – linearen und nichtlinearen – Modellen können durch das Problem des OR verzerrt werden (Cassel 2003; Bernstein et al. 2001; Voogt 2000; Atkeson 1999).

Insofern ist es nicht erstaunlich, dass die Diskussion über Gegenmaßnahmen zum OR die Forschungsagenda bis heute prägt. Dabei haben in Analogie zur theoretischen Ausrichtung und Begründung bis jetzt Zugänge überwogen, welche beim Messfehler angesetzt haben. In der Terminologie der Befragungsmethodik standen also so genannte Kontexteffekte im Vordergrund. Der Partizipationsfrage vorangestellte Fragen zu früheren oder anderen Teilnahmeforen könnten – so die Vermutung von Presser (1990) – den sozialen Druck abfedern bei der eigentlich interessierenden Frage, sich ins positive Licht zu rücken, bzw. die Verlegenheit über eine gelegentliche Nichtteilnahme zu reduzieren. Ein entsprechender experimenteller Test konnte die Annahme jedoch nicht bestätigen. Auch in den Experimenten von Abelson und seinem Team musste diese These des „two-time-frame procedure“ (Abelson et al. 1992) verworfen werden. Dennoch hatten diese Tests und insbesondere die Schlussfolgerungen große Ausstrahlung. Unverkennbar haben sie Belli und sein Team für ihre weiteren Untersuchungen inspiriert: Nach einem ersten erfolglosen Versuch (Belli et al. 1994) konnten sie denn auch die von Abelson et al. entwickelte Hypothese bestätigen: In zwei telefonischen „split-half“ Experimenten schnitt eine Formulierung, welche darauf abzielte, Probleme des source confusing zu beheben, besser ab als die ANES-Standardfrage und erwies sich besonders bei großen Zeitabständen zum befragten Ereignis als effektiv (Belli et al. 2001).

3 Analysen zum Overreporting im Kontext schweizerischer Wahl- und Abstimmungsstudien

Der schweizerische Kontext stellt ein ideales Forschungsfeld zur Überprüfung von verschiedenen OR-Thesen dar. Zwar liegen keine neueren Validitätsstudien vor, allein der Vergleich zwischen verschiedenen Typen von Urnengängen (Wahlen und Abstimmungen), zwischen Urnengängen mit ganz unterschiedlichen Partizipationsraten bei Sachabstimmungen ebenso wie die Zeitreihe für Sachabstimmungen erlauben es aber, einige Grundhypothesen zum OR zu testen. Darüber hinaus hatten wir die Möglichkeit, bei drei Untersuchungen je einen kleineren Methodentest anzuhängen. Auch wenn diese Tests aufgrund ihres Status als Zusatzuntersuchungen keineswegs hinreichend die Problematik zu erfassen vermögen, mit anderen Problemen der Datenqualität behaftet sein mögen (vgl. dazu Kapitel 3.2 und 4.) und die Hypothesen dem allgemeinen Design angepasst werden mussten, so erscheinen uns die Befunde dennoch berichterstattenswert und erfüllen hoffentlich die eine Funktion, weitere und bessere Tests anzuregen.

3.1 Die Hypothesen

Studien zur sozialen Erwünschtheit sowie Methodenuntersuchungen legen nahe, das telefonische Interview im Gegensatz zum persönlichen als eine Erhebungsmethode zu verstehen, die weniger Verzerrung provoziert. Mindestens auf aggregiertem Niveau lässt sich diese These anhand der Abstimmungs-Daten (VOX-Daten) testen, weil im Jahre 1995 die Umstellung auf dezentral durchgeführte, telefonische Interviews stattgefunden hat. Folglich lautet unsere Hypothese:

- 1.1. *Mit dem Übergang von persönlichen zu telefonischen Interviews wird der Durchschnittswert im Overreporting signifikant tiefer.*

Darüber hinaus haben wir die These der sozialen Erwünschtheit in einem Feldexperiment getestet. Wenn soziale Erwünschtheit das OR erklärt, dann sollte mit einem entsprechenden Wording der Frage das OR reduzierbar sein. Deshalb haben wir die Hypothese aufgestellt:

- 1.2. *Wenn im Einleitungstext zur Teilnahmefrage darauf hingewiesen wird, dass Nichtbeteiligung üblich und verbreitet ist, dann gibt es signifikant weniger Overreporting als bei einer Frage ohne diesen Hinweis.*

Abnehmende Parteibindungen und Partizipationsraten gelten gemeinhin als Indikator für Individualisierungsprozesse. Im Kontext der schweizerischen Referendumsdemokratie äußert sich dieser Prozess nicht in Form einer linearen Abnahme der Beteiligung, sondern in stark fluktuierenden Teilnahmequoten. Nur wenige Stimmbürger gehen aus staatsbür-

gerlichem Pflichtbewusstsein regelmäßig zur Urne, umso mehr dafür selektiv und je nach Betroffenheit. Daher formulieren wir die Hypothese:

1.3. Während die Stimmbeteiligung stark schwankt, nimmt das Overreporting im Verlaufe der Zeit ab.

In Bezug auf die Erinnerungsleistung und das source monitoring haben wir uns vor allem an den Begründungen von Abelson et al. (1992) sowie Belli et al. (1994 bzw. 1999) orientiert. So lässt sich das auf der individuellen Ebene formulierte Argument, wonach memory confusing vor allem bei Personen auftreten sollte, welche öfters teilnehmen (Abelson et al. 1992: 142), auf der Makro-Ebene des schweizerischen Kontexts wie folgt formulieren: Da sehr viel mehr nationale Abstimmungen als nationale Wahlen durchgeführt werden und erstere auch noch in recht enger Kadenz sowie mit mehreren Vorlagen erfolgen, sollte das source confusing vor allem bei Sachabstimmungen ein Problem darstellen. Entsprechend haben wir die Hypothese aufgestellt:

2.1. Bei Abstimmungen ist das Overreporting deutlich höher als bei Wahlen.

Für den experimentellen Test haben wir die Argumentation von Belli et al. (1999: 91) aufgenommen, wonach soziale Erwünschtheitsprobleme vor allem kombiniert mit Erinnerungsproblemen auftreten.

2.2. Wenn in einer Teilnahmefrage auch die Erinnerungsproblematik antizipiert wird, dann gibt es signifikant weniger Overreporting als bei einer Frage ohne diesen Hinweis.

Zudem haben wir in Anlehnung an Wright (1993) sowie Belli et al. (1999) den Faktor Zeit in Betracht gezogen. Je weiter weg das zu berichtende Ereignis – so die allgemeine Annahme dahinter –, umso stärker kommen response sets zum Tragen.¹⁹ Daher haben wir die gleiche Formulierung, die in der zweiten Befragung verwendet wurde, in einem dritten Test nochmals verwendet. Abgesehen vom anderen Thema und einem neuen Sampling hat sich dabei vor allem der Zeithorizont geändert: Die Frage wurde nun nicht unmittelbar nach der Abstimmung, sondern rund zwei Monate danach gestellt. Die Hypothese dazu lautete:

2.3. Der Effekt der Frageformulierung ist bei der Befragung mit einem größeren Zeitabstand größer als bei der Befragung unmittelbar nach dem Urnengang.

19 Es gilt aber zu beachten, dass das von Wright beobachtete response set ein Springen auf den Bandwagon beinhaltet, bei Belli et al. ist es das Overreporting.

Schließlich haben wir uns – soweit es die Daten zuließen – auch den Profilen gewidmet. Dabei sind wir aufgrund der sich abzeichnenden Tendenz (vgl. Abschnitt 2.2) und der theoretisch recht konsistenten Begründungen für das Experiment 1 von folgender Annahme ausgegangen:

- 3.1. *Auf das Wording zur Abfederung der sozialen Erwünschtheit reagieren vor allem die höher Gebildeten und politisch Interessierten/Involvierten mit weniger Overreporting.*

Im Gegensatz dazu sollte bei einem Wording, das den Vorschlägen von Belli et al. folgt, kein eindeutiges Profil erkennbar sein. Denn ein Wording, welches Erinnerungsprobleme und soziale Erwünschtheit kombiniert, kann zum Beispiel abnehmendes OR bei den weniger Gebildeten (Erinnerung) und höher Gebildeten (Erwünschtheit) gleichzeitig produzieren. Und ähnliches gilt, wenn man Gedächtnisprobleme als solche in Betracht zieht: Aufgrund der Theorie ist zu erwarten, dass reine Erinnerungsprobleme bei formal weniger Gebildeten und weniger Involvierten auftreten, während das source confusing jedoch bei formal höher Gebildeten und politisch Involvierten erwartet wird. Entsprechend vermuten wir:

- 3.2. *Auf ein Wording nach der Vorlage von Belli et al. reagieren Befragte an beiden Polen – höher Gebildete und stärker Involvierte ebenso wie weniger Gebildete und politisch kaum Interessierte – mit weniger Overreporting.*

3.2 Die Daten

Unsere Analysen basieren zum einen auf den „ex-post“ Analysen zu den eidgenössischen Urnengängen (VOX-Analysen) und Wahlen (Swiss Election Study) sowie auf drei „split-half“ Feldexperimenten.

VOX-Befragungen: Die Daten basieren auf einer zweistufigen (Haushalt – Personen) und geschichteten (Sprachregionen – Haushalte) ‚Zufallsstichprobe‘²⁰ von jeweils rund 1.000 Personen (ca. 700 Deutschschweiz, 240 Westschweiz, 60 italienische Schweiz). Die Befragungen wurden bis 1995 persönlich, ab Juni 1995 telefonisch, jedoch dezentral, durchgeführt. Die VOX sind vom Bund unterstützt, die Feldarbeit wird vom Meinungsforschungsinstitut GfS in Bern realisiert; die Universitäten Bern, Genf und Zürich zeich-

20 Die Auswahl der Personen erfolgt über die Geburtstagsmethode. Allerdings wird bei der GfS nicht alles dem Zufall überlassen, die Verantwortlichen der Stichprobe geben an: „Da sich auf der individuellen Ebene aber unterschiedlich starke Probleme mit der Erreichbarkeit ergeben, fügen wir zur Verbesserung der Stichprobenqualität Maximalgrößen bei“ (Technischer Bericht VOX Nr. 73: 13).

nen rotierend für Fragebogen und Analyse verantwortlich. Ab VOX Nr. 63 (Juni 1998) liegen Methodenberichte vor; die durchschnittlich berechnete Ausschöpfung beträgt 20.3 Prozent (Basis: bereinigte Bruttostichprobe, ohne stichprobenneutrale Ausfälle). Wir danken Lukas Golder von der GfS für die Kooperation.

Selects: Bis zu den Wahlen 1991 wurden die schweizerischen Wahlanalysen im Rahmen der VOX-Befragungen realisiert und folgten weitgehend diesen Stichproben. Seit 1995 werden die Nachwahlbefragungen von Selects (Swiss Electoral Studies), einem Zusammenschluss der Universitäten Bern, Genf und Zürich, organisiert. Basierend auf einer zweistufigen Zufallsstichprobe (Haushalt – Person), werden jeweils in den drei Wochen nach dem Wahlwochenende die telefonischen Befragungen durchgeführt (in der ganzen Schweiz, mit kantonalen Aufstockungen). 1995 wurden von der IPSO Sozial-, Marketing- und Personalforschung und vom Forschungsinstitut IHA-GfM zusammen über 7.000 Interviews realisiert. 1999 und 2003 war das Markt- und Sozialforschungsinstitut Link für die Befragung zuständig (3.258 resp. 5.891 Interviews). Die Ausschöpfung betrug 40.5, 51.3 und 46.9 Prozent (1995-2003).

Befragung 1: Das erste Feldexperiment wurde von DemoSCOPE zwischen dem 30. Januar und dem 5. Februar 2004 im Rahmen einer SNF-Studie über Informationsprozesse durchgeführt – also im Vorfeld der Abstimmungen über den Gegenvorschlag zur Avanti-Initiative, über die Verwahrungs-Initiative und über die Änderung des Mietrechts vom 8. Februar 2004. Die kurze Feldlaufzeit und das Interesse an Informationsprozessen von Teilnehmewilligen prägte die Stichprobe: Befragt wurden 1.004 Stimmberechtigte aus der Deutschschweiz; die Ausschöpfung der random/random-Stichprobe betrug 35.3 Prozent. Die Befragung fand vom zentralen Telefonlabor aus statt.

Befragung 2: Das zweite Feldexperiment wurde wiederum von DemoSCOPE eine Woche nach den Abstimmungen vom 16. Mai 2004 durchgeführt, bei denen es um folgende Vorlagen ging: 11. AHV-Revision, Erhöhung der Mehrwertsteuer zugunsten von AHV/IV und Steuerpaket. Hier wurden 700 Personen in der Deutschschweiz in einem random/quota-Verfahren befragt. Das Total der verwendeten Ausgangsadressen betrug 3.472; 837 der Ausfälle waren neutral, 1.935 könnten verzerrend wirken (1.095 nicht erreichbar; 839 Verweigerungen); womit sich eine ‚Ausschöpfung‘ von 26.6 Prozent ergab.

Wir danken Frau Tschurennev von DemoSCOPE für die Begleitung.

Befragung 3: Das dritte Feldexperiment behandelte wiederum die Abstimmung vom 8. Februar 2004, wurde jedoch 77 Tage nach der Abstimmung realisiert. Es handelte sich um eine Methodenstudie (Unit-Nonresponse) im Auftrag des SIDOS und in Kooperation mit dem Bundesamt für Statistik, welches für das Sampling zuständig war. Bei Subsample 1

wurde nach der ersten Verweigerung ein Incentive verschickt und es wurden Konvertierungsinterviews durchgeführt. Für Subsample 2 fanden Konvertierungsinterviews ohne Incentive statt; Subsample 3 wurde gemäß üblichem Standard, wo sofort die Zielperson ausgewählt wird, durchgeführt. Aufgrund der Testanlage wurden zwei Kontakte hergestellt: zwischen dem 17.3.-3.4. (Haushalt) und vom 26.4.-15.5.2004 für die eigentlichen Personeninterviews: In diese Phase fiel somit auch das Experiment. Vom ursprünglichen Adressenset (750) konnten 369 Interviews realisiert werden; womit sich nach Abzug der stichprobenneutralen Ausfälle auf der Ebene von Haushalt und Personen ($n = 86$) eine Ausschöpfung von 55.6 Prozent ergibt. Wir danken dem Studienleiter, Dominique Joye, für den großzügigen Zugang zu dieser Erhebung.

Die Daten stammen aus Erhebungen, die – mit einer Ausnahme – für andere Zwecke realisiert wurden. Daher müssen wir insbesondere auch bei den unabhängigen Variablen auf unterschiedliche Operationalisierungen zurückgreifen. Das zweite und dritte Feldexperiment wurde fast zeitgleich realisiert; die Nummerierung folgt der Argumentationslogik in diesem Artikel. Über die genaue Frageformulierung gibt Tabelle 1 Auskunft. Die Zuweisung zu den Experimentalgruppen war in allen Fällen randomisiert. Der Einfachheit halber sprechen wir von Experimentalgruppe dort, wo ein gezieltes Wording vorgenommen wurde und von Kontrollgruppe dort, wo eine Standardfrage verwendet wurde.

Tabelle 1 Frage- und Antwortformulierungen der Feldexperimente

<p>1. Befragung Experimentalgruppe Bei solchen Abstimmungen geht normalerweise mehr als die Hälfte der Stimmberechtigten nicht an die Urne, es gibt schließlich noch anderes als Politik.</p> <p>Wie ist das bei Ihnen, nehmen Sie voraussichtlich an der Abstimmung vom 8.2. teil oder nicht, oder haben Sie sich noch nicht entschieden? ja, habe schon abgestimmt ja, will abstimmen gehen nein, will nicht abstimmen gehen ich habe noch nicht entschieden</p> <p>weiß nicht keine Angabe</p>		<p>Kontrollgruppe (Standardfrage)</p> <p>Wie ist das bei Ihnen, nehmen Sie voraussichtlich an der Abstimmung vom 8.2. teil oder nicht, oder haben Sie sich noch nicht entschieden? ja, habe schon abgestimmt ja, will abstimmen gehen nein, will nicht abstimmen gehen ich habe noch nicht entschieden</p> <p>weiß nicht keine Angabe</p>	
<p>2. Befragung Experimentalgruppe</p>	<p>3. Befragung Experimentalgruppe</p>	<p>2. Befragung Kontrollgruppe (Standardfrage)</p>	<p>3. Befragung Kontrollgruppe (Standardfrage)</p>
<p>Am Wochenende fand eine Abstimmung über das Steuerpaket, die 11. AHV-Revision und die Erhöhung der Mehrwertsteuer zu Gunsten der AHV/IV statt.</p> <p>Versuchen Sie sich nochmals kurz zu erinnern. Haben Sie an der Abstimmung vom 16. Mai teilgenommen, wollten sie ursprünglich teilnehmen, haben sie nicht teilgenommen oder können Sie sich nicht mehr erinnern?</p> <p>Ja, habe teilgenommen</p> <p>Wollte ursprünglich teilnehmen, war dann aber verhindert</p> <p>Wollte ursprünglich teilnehmen, aber konnte mich nicht entscheiden</p> <p>Nein, habe nicht teilgenommen</p> <p>(Nicht vorgelesen): Weiß nicht mehr Nicht stimmberechtigt</p>	<p>Vor etwa knapp zwei Monaten, am 8. Februar fand eine eidgenössische Volksabstimmung über die Verwahrungsinitiative, den Avanti-Gegenvorschlag und eine Änderung des Mietrechts statt.</p> <p>Versuchen Sie sich nochmals kurz an diese Zeit zu erinnern. Haben Sie an der Abstimmung vom 8. Februar teilgenommen, wollten sie ursprünglich teilnehmen, haben sie nicht teilgenommen oder können Sie sich nicht mehr erinnern?</p> <p>Ja, habe teilgenommen</p> <p>Wollte ursprünglich teilnehmen, war dann aber verhindert</p> <p>Wollte ursprünglich teilnehmen, aber konnte mich nicht entscheiden</p> <p>Nein, habe nicht teilgenommen</p> <p>(Nicht vorgelesen): Weiß nicht mehr</p>	<p>Am Wochenende fand eine Abstimmung über das Steuerpaket, die 11. AHV-Revision und die Erhöhung der Mehrwertsteuer zu Gunsten der AHV/IV statt.</p> <p>Wie war das bei Ihnen, haben Sie an der Abstimmung vom 16. Mai teilgenommen oder nicht?</p> <p>ja nein (Nicht vorgelesen): keine Angabe</p>	<p>Vor etwa knapp zwei Monaten, am 8. Februar, fand eine eidgenössische Volksabstimmung über die Verwahrungsinitiative, den Avanti-Gegenvorschlag und eine Änderung des Mietrechts statt.</p> <p>Wie war das bei Ihnen, haben Sie an der Abstimmung vom 8. Februar teilgenommen oder nicht?</p> <p>ja nein (Nicht vorgelesen): keine Angabe</p>

3.3 Befunde

Tabelle 2 Datensätze und Overreporting im Überblick – Angaben in Prozent

Datensatz	n	Selbst-deklarierte Teilnahme/Absicht ²⁾	Offizielle Teilnahme	Overreporting (d)
66 Volksabstimmungen (VOX-Befragungen 1977-2004) ¹⁾	ca. 1.003	58.1	42.6	15.5
7 Wahlen (VOX-Befragungen und Selects 1995 ff.) ²⁾	ca. 2.850	63.5	47.6	15.9
1. Befragung vor der Abstimmung vom 8.2.2004	1.000	86.5 ⁴⁾	45.0	41.5
2. Befragung nach der Abstimmung vom 16.5.2004	700	75.9	50.2	25.7
3. Befragung 2 Monate nach der Abstimmung vom 8.2.2004	369	67.1	45.0	22.1

¹⁾ Es handelt sich um Durchschnittswerte basierend auf den substanzialen und gerundeten Zahlen für jede VOX

²⁾ Durchschnittswerte der Nachwahlbefragungen 1971-2003, ohne 1983

³⁾ in den nachfolgenden Tabellen wird der Einfachheit halber von Teilnahme gesprochen

⁴⁾ inkl. „ja, ich habe die Absicht teilzunehmen“

Als erstes zeigt sich in Tabelle 2, dass das Phänomen des OR in allen verwendeten Datensätzen zu beobachten ist, die Vorwahlbefragung aber ähnlich wie schon von Jackman (1999: 18) aufgezeigt, besonders anfällig dafür ist. Bei den „ex-post“ Analysen zu den eigenössischen Urnengängen ist das OR zwar vergleichsweise gering, bewegt sich aber ziemlich genau auf dem Niveau, das in den Wahlstudien anderer Länder zu beobachten ist.

Wenn wir gemäß Hypothese 1.1. den Blick etwas stärker auf diese Abstimmungserhebungen fokussieren, dann zeigt sich zudem: Hinter dem Mittelwert 15.5 versteckt sich eine hohe Spannweite von 17 Prozentpunkten. Das OR kann bis auf acht Prozentpunkte sinken und bis auf 25 Prozentpunkte hinaufklettern. Ein abnehmender Trend ist jedoch nicht zu beobachten und Tabelle 3 illustriert, dass der Übergang von persönlichen (face-to-face) zu telefonischen Interviews keine signifikante Reduktion mit sich gebracht hat. Verknüpft mit der Tatsache, dass wir es nie mit einem Underreporting zu tun haben, kann man aus diesen Daten noch keinen Beweis gegen die soziale Erwünschtheitstheorie ablesen. Vielleicht haben wir auch einfach einen Hinweis dafür, dass es gerade im Rahmen von schnell durchgeführten Telefonbefragungen schwierig ist, dem Phänomen mit einer Maßnahme auf der Ebene des Wording zu begegnen.

Tabelle 3 Overreporting und Erhebungsmethode (Interviewstil) in den VOX-Analysen

	Mittelwert Face to face	Mittelwert telefonisch	Quadrat- summe	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	15.4	15.4	0.005	1	0.005	0.001	0.982

Mit Hilfe des „split-half“ Experiments in der 1. Befragung konnten wir die These mit einem Zugang von höherer interner Validität testen. Ein Einleitungstext zur Abfederung der sozialen Erwünschtheit sollte das OR in der Experimentalgruppe signifikant verringern. Das in Tabelle 4 präsentierte Datenmaterial macht allerdings deutlich, dass die Nullhypothese zur Hypothese 1.2. nicht widerlegt werden kann. Damit bestätigen sich die Befunde aus den USA. Eine leicht veränderte andere Formulierung mit Antizipation der sozialen Erwünschtheit vermochte weder auf aggregiertem Niveau der ANES noch in experimentellen Wording-Tests einen Effekt zu produzieren, wobei wir bei der vorliegenden Untersuchungsanlage über die Gründe nur spekulieren können. Probleme der Vorwahlbefragung und des Sampling könnten kombiniert mit der Tatsache, dass der experimentelle Stimulus im Rahmen eines telefonischen Interviews schlicht zu schwach ist, dafür verantwortlich sein.

Tabelle 4 Effekt des Wording (soziale Erwünschtheit) auf das Antwortverhalten

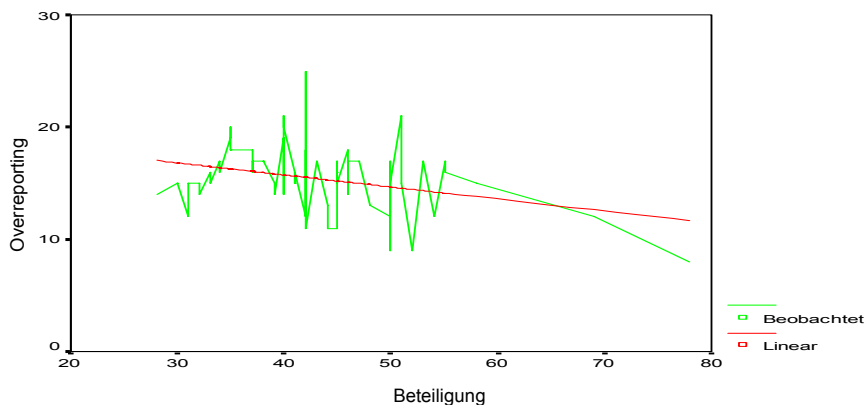
Test	Treatment	Teilnahme ja ¹⁾ in Prozent	Teilnahme nein in Prozent	n	Signifikanz ²⁾
1. Befragung vor der Abstimmung Februar	EG	91.0 (436)	9.0 (43)	479	0.475
	KG	90.7 (429)	9.3 (44)	473	

Lesebeispiel: Bei der ersten Befragung haben in der Experimentalgruppe (EG) mit dem SD-Wording (vgl. Tabelle 1) 91.0 Prozent angegeben teilzunehmen, in der Kontrollgruppe mit der Standardfrage waren es 90.7 Prozent. Der Unterschied zwischen den Gruppen ist somit nicht signifikant.

¹⁾ inkl. Teilnahmeabsicht.

²⁾ Exakter Fisher-Test, einseitig.

Auch die Hypothese 1.3. lässt sich nicht bestätigen. Die Beteiligung an Abstimmungen ist zwar in der Tat stark schwankend und die Kurve dokumentiert den selektiven Umrang in der Schweiz. Das Overreporting nimmt aber im Verlaufe der Zeit keineswegs ab ($Rsq = 0.002$; $df = 64$; $F = 0.13$; $Sig. = 0.724$; $b1 = -0.000$). Allerdings impliziert die ausbleibende Abnahme auch nicht, dass die konträre These von Burden (2003 und 2000) zutrifft, wonach das OR im Verlaufe der Zeit sogar zunimmt. Hingegen stellen wir einen schwachen, aber eindeutig negativen Zusammenhang zwischen der Stimmbeteiligung und dem OR fest ($Rsq = 0.102$; $df = 64$; $F = 7.27$; $Sig. = 0.009$; $b1 = -0.107$).

Grafik 1 Overreporting nach Stimmbeteiligung

Insofern können wir Granberg & Holmberg (1991: 457) bestätigen, welche darauf hinweisen, dass die Validität der Partizipationsfrage eine Funktion der hohen Teilnehmerate darstellt. Das wiederum deutet – im Sinne eines plausiblen, aber keineswegs kausalen Arguments – darauf hin, dass wir es bei der VOX-Analyse weniger mit einem Mess- sondern vielmehr mit einem Sampling-Problem zu tun haben. Im Moment, wo die „population at risk“ – die Nichtwähler – in der Population weniger stark vertreten ist, scheint auch deren Unterrepräsentanz weniger problematisch.

Aus theoretischer Sicht ist diese Beobachtung eines negativen Zusammenhangs allerdings schwierig zu erklären. Im Fall eines positiven Zusammenhangs könnte man mit Rückgriff auf die soziale Erwünschtheit behaupten, dass bei einer höheren Stimmbeteiligung das OR steigt, weil Befragte – sofern sie das Wissen über die Teilnahme haben – mit einer Art Konsensus-Heuristik nach dem Motto „wenn so viele gingen, dann muss das wohl ein wichtiger Urnengang gewesen sein“ reagieren. In etwa so argumentieren auch Karp & Brockington (2005) und stützen ihre These mit gepoolten Individualdaten aus Validitätsstudien von fünf Ländern. Für das hier vorliegende Muster mit sinkendem OR lässt sich indessen mit der Erwünschtheitstheese kaum eine Erklärung konstruieren. Somit werden bei diesem Analyseschritt nicht nur die Grenzen eines Zugangs mit Aggregatdaten offenkundig, sondern es wird deutlich, dass Mess- und Sampling-Probleme gleichzeitig untersucht werden sollten.

Ausgehend von den Ausführungen zum source monitoring haben wir die Hypothese aufgestellt, bei den schweizerischen Abstimmungsanalysen müsse das OR ausgeprägter sein als bei den Wahlen. Ein Vergleich der entsprechenden Mittelwerte vermag diese Annahme

nicht zu bestätigen (vgl. Tabelle 5). Werden sämtliche vorliegenden Studien miteinander verglichen, unterscheiden sich die Mittelwerte nicht (15.4 Abstimmungen; 15.8 Wahlen). Bei einer Reduktion auf jene Fälle, die mit vergleichbaren Sampling-Prozeduren erstellt wurden (ab 1995) stellen wir das Gegenteil der Hypothese fest. Das OR ist bei den Wahlen ausgeprägter. Aufgrund dieses einfachen bivariaten Zugangs und der Tatsache, dass wir das source monitoring nicht direkt erfassen können, ist das noch kein starker Gegenbeweis zur These. Aber er lädt zumindest dazu ein, sich über die Bedingungen des source confusing und möglicher Gedächtniseffekte Gedanken zu machen. Unsere Annahme war, dass Abstimmungen aufgrund der hohen Frequenz anfällig für OR sind. Die vorliegenden Daten legen indessen nahe, davon auszugehen, dass – wenn dieser kognitive Faktor überhaupt eine Rolle spielt – nicht Häufigkeit mit Konfusion einhergeht und Konformität produziert, sondern eher Seltenheit oder geringer Stellenwert mit Vergesslichkeit einhergeht und Konformität zur Folge hat. Damit hat vielleicht Presser (1984: 353) mit seiner Vermutung Recht behalten: „But the process may also have been a more subtle one, affecting respondents’ memories. This seems especially likely for an action like voting [gemeint ist das Wählen], which is neither central to the lives of most people nor performed frequently.”²¹

Tabelle 5 Overreporting nach Typ des Urnengangs (alle und ab 1995)

	Mittelwert Abstimmungen	Mittelwert Wahlen	Quadrat- summe	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen (alle)	15.4	15.8	1.261	1	1.261	0.082	0.775
Zwischen den Gruppen (ab 1995)	15.4	20.2	64.24	1	64.24	6.259	0.018

Ohnehin stößt man aber bei einem kognitionspsychologischen Konstrukt mit Aggregatdaten an analytische Grenzen, weshalb wir uns lieber den Individualdaten widmen wollen. Für den Test der Hypothesen 2.2. und 2.3. haben wir in der 2. und 3. Befragung zwei Feldexperimente geschaltet, die sich – abgesehen vom sicher nicht zu vernachlässigenden thematischen Wechsel und dem unterschiedlichen Sampling (inkl. Ausschöpfungsquote) – im Zeitfaktor unterscheiden: Das zweite Experiment wurde unmittelbar nach der Mai-Abstimmung realisiert; das dritte mehr als zwei Monate nach der Februar-Abstimmung.

21 In der Tat teilen die befragten Stimmberechtigten in den VOX-Analysen immer wieder mit, Abstimmungen seien wichtiger als Wahlen.

Bei der Befragung nach den Abstimmungen im Mai hat das unterschiedliche question wording keinen Effekt; die beiden Splits unterscheiden sich nicht signifikant (vgl. Tabelle 6). Bei der Befragung nach den Abstimmungen im Februar hingegen, die erst zwei Monate später stattgefunden hat, ist der erwartete Effekt zu beobachten: Mit der Experimentalfrage, welche die Erinnerungsproblematik (und die soziale Erwünschtheit) berücksichtigt, geben weniger Befragte an, dass sie an den Abstimmungen teilgenommen haben ($p = 0.088$). Ein Wording-Effekt in einer kurzfristig angelegten Befragung ist somit nicht effektiv; bei Hypothese 2.2. kann die Nullhypothese nicht verworfen werden. Hingegen haben wir – auch wenn der direkte Vergleich zwischen den zwei Befragungen aufgrund anderer Rahmenbedingungen nur beschränkt möglich ist – ein starkes Indiz für Hypothese 2.3. Liegt das Ereignis weiter zurück, so kann mit einem entsprechenden Wording ein positiver Effekt erzielt werden; das Overreporting nimmt ab.

Tabelle 6 Effekt des Wording (source monitoring) auf das Antwortverhalten

Test	Treatment	Teilnahme ja in Prozent	Teilnahme nein ¹⁾ in Prozent	n	Signifikanz ²⁾
2. Befragung nach der Abstimmung vom 16. Mai	EG	77.4 (270)	22.6 (79)	349	0.219
	KG	74.6 (261)	25.4 (89)	350	
3. Befragung 2 Monate nach der Abstimmung vom 8. Februar	EG	63.9 (108)	36.1 (61)	169	0.088
	KG	71.7 (104)	28.3 (41)	145	

¹⁾ inkl. „wollte ursprünglich teilnehmen, war dann aber verhindert“ bzw. „konnte mich aber nicht entscheiden“

²⁾ Exakter Fisher-Test, einseitig

Die Wirkung scheint jedoch nicht überragend.²² Das Overreporting bleibt selbst in der Experimentalgruppe bestehen (63.9 Prozent selbst-deklarierte Teilnehmende gegenüber 45 Prozent in der offiziellen Statistik). Zudem ist die provozierte Prozentpunkt-Differenz zwischen der Experimental- und der Kontrollgruppe mit 7.8 nicht sehr ausgeprägt. Denn im Vergleich dazu wird in unserer Referenzstudie bei einem Design mit kleinerem time lag (56 Tage) eine starke Differenz von 17.6 Prozentpunkten provoziert (Belli et al. 1999: 102). Während dort in der Kontrollgruppe mit der ANES-Standardfrage immer noch 74.9 Prozent angaben, sie hätten an den Wahlen teilgenommen, sinkt der Anteil in der Experimentalgruppe auf 57.3 Prozent. Eine Erklärung für diesen starken Effekt liegt vermutlich im Wording des Experiments:

22 Der Begriff „scheint“ ist hier durchaus am Platz, weil wir Effekte des Sampling, das vielleicht eine höhereeteiligungsrate erklärt, nicht kontrollieren können und keine Validierung der Daten möglich ist.

Tabelle 7 Wording im Experiment von Belli et al. (1999: 106)

Experimental Wording, National Study	
<i>The next question is about the elections in November. In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time.²³ We also sometimes find that people who thought that they had voted actually did not vote. <u>Also, people who usually vote may have trouble saying for sure whether they voted in a particular election. In a moment I am going to ask you whether you voted on Tuesday, November 5th, which was {time fill} ago.</u> Before you answer, think of a number of different things that will likely come to mind if you actually did vote this last past election day; things like whether you walked, drove, or were driven by another person to your polling place {pause}, what the weather was like on the way {pause}, the time of day that was {pause}, and people you went with, saw, or met while there {pause}. <u>After thinking about it, you may realize that you did not vote in this particular election {pause}.</u> Now that you've thought about it, which of these statements best describes you?"</i>	
I did not vote in the November 5 th election (1);	I am sure I voted in the November 5 th election (4)
I thought about voting this time but didn't (2);	I voted by absentee ballot (volunteered, 7).
I usually vote but didn't this time (3)	

Während diese Formulierung darauf abzielt, soziale Erwünschtheit zu reduzieren (kursiv) und vor allem das Retrievement zu erleichtern (unterstrichen), indem mit ‚kognitiven Bildern‘ die relevanten Informationen im Gedächtnis verfügbar gemacht werden, könnte diese Formulierung allerdings auch eine Art Präzisions- oder Kontrolleffekt auslösen (Formulierungen ohne Hervorhebung). Wenn wir nämlich in Betracht ziehen, dass die Befragten von Vorgaben in einem Fragebogen ‚lernen‘ (Schwarz & Scheuring 1992), dann sind mindestens zwei weitere Erklärungen für die Effektivität des von Belli et al. gewählten, aber, zumindest in telefonischen Interviews ungeeigneten Wordings möglich: Zum einen könnte es sein, dass mit der ausführlichen Frage den Befragten eindringlich genug kommuniziert wird, dass hier nun große Sorgfalt verlangt ist. Dann hätten wir es vielleicht mit jenem Präzisionseffekt zu tun, auf den bereits eine Koautorin von Abelson aufmerksam gemacht hat (Loftus et al. 1990). Zum anderen ist nicht auszuschließen, dass die Befragten nach dieser langen Einleitung vermuten, die Frage, wie man zum Wahllokal gekommen sei, werde später im Interview nochmals aufgenommen. Oder die Konkretheit der Situationsschilderung vermittelt den Interviewten sogar der Eindruck, man hätte Zugang zu Kontrolldaten (wenn auch nur Freunde, die die Wählenden vermeintlich begleitet haben). Vielleicht haben also Belli und sein Team ein Treatment entwickelt, das – unbeabsichtigt aber effektiv – einen Gedanken aufnimmt, den Abelson et al. (1992: 151) formuliert haben und den klassischen „bogus pipeline studies“ entnommen ist (Jones & Sigall 1971). Demnach können Antwortverzerrungen durch soziale Erwünschtheit vor

23 Der Text bis hierhin entspricht der ANES-Standardfrage.

allem dann reduziert werden, wenn die Befragten meinen, unaufrichtige Antworten könnten durch Kontrollen entdeckt werden (Tourangeau et al. 2000: 277).²⁴

Abschließend möchten wir die zentralen Befunde zu den Profilen jener zusammenfassen, die auf ein Treatment zur sozialen Erwünschtheit und/oder zu Gedächtniseffekten reagieren (vgl. Tabelle 8).

Tabelle 8 Subgruppenanalyse und Merkmale mit Effekt

Merkmal/Subgruppe	Wo/ Nummer der Befragung	Effekte: Differenz in Prozentpunkten	
		Wie/ Abnahme Ja Anteil	Wie/ Zunahme Ja Anteil
Mittleres Alter (35- bis 54-Jährige)	3	14.2* (117)	
<i>Bildung</i>			
Primar- und Sekundarschule	2	14.6 * (110)	
Sekundar-, Real-, Bezirksschule	2	17.5 * (87)	
Berufsschule	3 / 2	19.1*** (177)	7.3* (337)
Alles außer Primar- und Sekundarschule	3	12.8** (254)	
Alles außer Uni/ETH	3	8.1* (287)	
Berufsschule bis Uni	2		5.8* (583)
Teilzeitarbeit	3	17.2* (64)	
Vertrauen in PolitikerInnen			
Wenig (sehr wenig und wenig)	3	13.9* (128)	
Sehr wenig	3	23.9* (45)	
<i>Interesse für Politik</i>			
kaum	3	23.1** (74)	
Gar nicht	1		43.2 ** (35)
<i>Wahrnehmung Politik als kompliziert</i>			
ziemlich häufig	3	29.2** (57)	
häufig (ziemlich häufig und häufig)	3	23.6 *** (106)	
Effektivität: wenig Politiker kümmern sich	3	12.7* (120)	
Keine Petition unterzeichnet	3	14.9** (208)	
Nie Diskussionen über Politik	1		26.7** (65)
Beitritt in Partei vorstellbar	2		22.7* (110)
Debatten über Abstimmung verfolgt	2		9.2 * (226)
Stichprobe 1: mit Incentives	3	22.2 (96) **	

Lesebeispiel: Bei der Variable ‚Alter‘ konnte einzig unter den 35- bis 54-Jährigen eine signifikante Reaktion auf das Wording registriert werden – und zwar in Befragung 3: Während in der ‚Kontrollgruppe‘ 73.6 Prozent angeben, sie hätten teilgenommen, sind es in der Experimentalgruppe 59.4 Prozent. Somit ergibt sich die in der Tabelle ausgewiesene Differenz von 14.2 Prozentpunkten. In die entsprechende Vierfeldertabelle sind insgesamt 117 Fälle eingeflossen.

Signifikanz: Exakter Fisher-Test, 1-seitig: *** $p <= .01$, ** $p <= .05$, * $p <= .1$

24 Entsprechend metaphorisch ist der Begriff „bogus pipeline“. Die Experimentierer vermitteln den Befragten den Eindruck, sie hätten Zugang (eine Pipeline) zu den ‚wahren‘ Reaktionen bzw. Verhaltensweisen.

In Bezug auf die erste Befragung und die soziale Erwünschtheitstheorie bestätigt sich dabei, dass der schwache Nettoeffekt keineswegs durch Bewegungen in unterschiedliche Richtungen in den Subgruppen zustande kommt, sondern in der Tat schwach ist. Darüber hinaus, und das interessiert hier mehr, zeigt sich, dass die Empirie in doppelter Hinsicht gegen Hypothese 3.1. spricht: Erstens stellen wir nicht bei jenen Personen, die den aufrichtigen Wählern gleichen, eine Reaktion fest, sondern bei jenen Befragten, die sich offen als „gar nicht interessiert“ bezeichnen und nie über Politik diskutieren. Zweitens löst das Treatment das Gegenteil dessen aus, was beabsichtigt ist. Bei den Stimmberechtigten, die mit Politik gar nichts anfangen können, nimmt das OR bei Verwendung der Erwünschtheitsfrage zu.²⁵ Wir sind hier wohl mit einem Mechanismus konfrontiert, der im Zusammenhang mit den „two-time frame“ Experimenten auch schon in Betracht gezogen wurde (Abelson et al. 1992: 142): Erst über das Wording scheint man hier auf die soziale Erwünschtheit der Stimmbeteiligung aufmerksam gemacht. Oder der explizite Hinweis auf verschiedene Gruppen von Wählern löst ein Distinktionsbedürfnis aus, ähnlich wie es auch dem Third-Person-Effekt inhärent ist. Damit löst die Frageformulierung nicht ein Problem, sondern schafft eines.

Und als wäre das der Komplexität nicht genug, stellen wir bei Befragung 2 fest, dass wir hier ähnliche negative Effekte des Wording haben, diesmal jedoch nicht die schlechter Integrierten, sondern Personen mit einem höheren Bildungsniveau (ab Berufsschule aufwärts) und politisch Involvierte (Beitritt in Partei vorstellbar, Debatten über Abstimmung verfolgt) noch mehr overreporten. Hypothese 3.2. ist somit dadurch widerlegt, dass die Bewegungen nicht wie vermutet in Richtung Abnahme des OR weisen. Hinzu kommt, dass wir bei Befragung 3 in der Tendenz und entgegen der Annahme ein Profil erkennen. Bei den soziodemographischen Variablen sind es eher die mittleren Kategorien, die reagieren. Bei den politischen Variablen sind es vor allem die Nichtinteressierten, die auf das Wording reagieren: Jene, die noch nie eine Petition unterzeichnet haben; jene, die glauben, sehr wenig Politiker würden sich darum kümmern, was die Wählenden denken; und Befragte, die nur mit zusätzlichen Incentives zur Teilnahme an der Befragung bewegt werden konnten. Dabei könnte die Reaktion der mittleren Kategorien bei den soziodemographischen Variablen durchaus in Einklang mit der source confusing stehen und impliziert vielleicht, die Annahme umzuformulieren: Gerade jene, die weder besonders stark noch besonders schwach ausgestattet sind, unterliegen einem source confusing. Insgesamt können wir aber vor allem festhalten: Der konsistent uniforme Effekt und die

25 Während unter der einfachen Teilnahmefrage ohne Einleitungssatz nur 23.5 Prozent der Befragten angeben, sie hätten teilgenommen, sind es unter der Bedingung des gezielten Wordings 66.7 Prozent, womit die in der Tabelle ausgewiesenen 43.2 Prozentpunkte Differenz zustande kommen. Der Chi-Quadratwert mit Yates Korrektur bestätigt die Signifikanz ($p = 0.026$).

vergleichsweise großen Prozentpunkt-Differenzen im dritten Feldexperiment unterstreichen die Annahme von Hypothese 2.3, dass, wenn das Ereignis weit zurückliegt, ein starker Wording-Effekt nachzuweisen ist. Welches die Ursachen dafür sind, bleibt aber solange unklar, bis Probleme der sozialen Erwünschtheit, der Erinnerung oder des source confusing nicht auch direkt erfasst werden.

4 Fazit

Das Overreporting politischer Partizipation – so zeigt die vorliegende Literaturdurchsicht ebenso wie das Länderbeispiel Schweiz – stellt ein offenkundiges Problem dar. Es wird in allen bisher untersuchten und etablierten Demokratien festgestellt und scheint nicht nur pankulturell, sondern auch auf hohem Niveau verbreitet zu sein. Auf aggregiertem Niveau werden leicht Differenzen von 15 Prozentpunkten erreicht und der Anteil in Validitätsstudien ‚entdeckter‘ unaufrichtiger Wähler kann bald 25 Prozent erreichen.

Dabei ist das Problem auch deshalb nicht zu unterschätzen, weil es mittlerweile zahlreiche Hinweise dafür gibt, dass das Misreporting in Befragungen zu einem „Overreporting der Sozialwissenschaftler“ führen kann. Wenn Overreporters überdurchschnittlich stark politisch interessiert sind, verleiten sie die Sozialwissenschaftler dazu, den Zusammenhang zwischen Interesse und Teilnahme in der Politik zu überschätzen. Unter Umständen bewegen wir uns also in einer Spirale des Common Sense, die uns nicht gefallen kann. Die einen orientieren ihre Antworten, die anderen ihre wissenschaftliche Modelle daran (Presser & Traugott 1992: 86). Zudem liefert die Analyse der VOX-Nachbefragungen einen weiteren Hinweis dafür, dass mit der Verbreitung der Telefoninterviews²⁶ entgegen der vielfach geäußerten Hoffnung nicht automatisch eine Abnahme des Overreporting einhergeht. Es scheint,²⁷ dass die vergleichsweise anonyme und unverbindliche Erhebungsform die starke Verankerung der gesellschaftlichen Norm „Abstimmen ist Bürgerpflicht“ nicht zu durchbrechen vermag. Umgekehrt kann es sehr gut sein, dass diese Erhebungsmethode Maßnahmen auf der Ebene des Wording erschwert. Die Tatsache jedenfalls, dass unser erstes Feldexperiment keine Wirkung und das Experiment von Belli und seinem Team mit einem sehr langen Einleitungstext dagegen viel Wirkung zeigte, kann auch damit zusammenhängen: Vielleicht verhindert die schnelle und spontane telefonische Erhebungsform, dass das Wording der Frageformulierung – in das die Forscher so viel investieren – in seiner Feinheit über-

26 Gemäß Statistik des Verbandes Schweizer Marketing- und Sozialforscher (SMS) wurden im Jahr 2000 66 Prozent der Befragungen telefonisch durchgeführt; auch in Deutschland dominiert diese Erhebungsmethode zu diesem Zeitpunkt (Gabler & Häder 2002).

27 Die zurückhaltende Formulierung ist angezeigt, weil Sampling-Effekte nicht ausgeschlossen werden können.

haupt wahrgenommen wird. Erst eine Frageformulierung, von der eigentlich jedes Lehrbuch aufgrund der Länge abrät, vermag vielleicht die notwendige Aufmerksamkeit der Befragten zu provozieren.

Insgesamt können also diese Maßnahmen nicht wirklich überzeugen. Zum ersten deshalb, weil sie zumeist Sampling und Unit-nonresponse-Probleme vernachlässigen, obwohl wir seit Brehm (1993) und den Studien von Voogt & Saris (2003) wissen, dass politische Partizipation und Befragungspartizipation korrelieren bzw. Sampling- und Messfehler in jeder Befragung in Kombination wirken können (Jackman 1999: 14). Hinweise für die Bedeutung der Sampling-Dimension konnten wir – oder mussten wir trotz Ermangelung entsprechender Daten – im Rahmen der oben stehenden Ausführungen mehrmals andeuten. Deshalb haben wir für die vielfach genutzten VOX-Analysen auch einen Strukturdatenvergleich vorgenommen. Die Daten verdeutlichen, dass die Stichproben in die erwartete Richtung verzerrt sind: potenzielle WählerInnen sind übervertreten (Angehörige der deutschen Schweiz, ältere Personen und höher Gebildete). Wir haben also auch hier starke Hinweise dafür, dass Overreporting auch auf Sampling-Aspekte zurückzuführen ist.

Unbefriedigend sind die besprochenen Maßnahmen zum zweiten auch deshalb, weil sie sich als riskantes Unterfangen entlarven können. Zuweilen scheinen sie nicht den Druck der Anpassung an die soziale Norm abzubauen, sondern lediglich diese soziale Norm (wieder) in Erinnerung zu rufen. Darüber hinaus sind die theoretischen Annahmen noch zu wenig klar und wir kennen die kausalen Prozesse zu schlecht, um experimentelle Manipulationen wirklich legitimieren zu können. Deshalb führt unseres Erachtens kein Weg daran vorbei, die möglichen Maßnahmen gegen ein Overreporting in zweierlei Richtungen weiterzudenken.²⁸

Erstens sollten die erklärenden Faktoren (soziale Erwünschtheit, Erinnerungsprobleme und source confusing) in Studien zum Overreporting parallel erfasst werden.²⁹ Denn nur mit einer solchen Parallelerhebung lässt sich klären, ob soziale Erwünschtheit (immer noch)³⁰ ein klares soziodemographisches Profil hat oder stärker vom Befragungskontext

28 Auf die mögliche Maßnahme des „random response“ wird hier nicht eingegangen, weil sie unseres Wissens vor allem in persönlichen Interviews eingesetzt wird, hier aber telefonische Befragungen im Vordergrund stehen. Für mögliche Testanlagen vgl. Locander et al. (1976: 270). Eine weitere Maßnahme würde darin bestehen, dass die kommerziellen Institute ihre Kontrollfragen und Prozeduren für die Erstellung von genauen Prognosen bekannt geben würden. Das kommt vielleicht einem frommen Wunsch gleich, und bis eine solche Studie realisiert ist, sei auf ein ähnliches Vorgehen kombiniert mit einer Validitätsstudie verwiesen: Miller (1952).

29 Zumindest in Bezug auf die soziale Erwünschtheit liegen ja mittlerweile auch deutschsprachige und ‚vergleichsweise‘ kurze Fragebatterien vor (Krebs 1991; Stöber 1999).

30 Zumindest ist das die Einschätzung von Krebs (1991: 7).

beeinflusst ist und damit als volatiler, aber messbarer Persönlichkeitszug zu verstehen ist. Zudem ließen sich Maße einer SD-Skala als Kontrollvariable in Modellschätzungen einsetzen, um zu demonstrieren, dass berechnete Zusammenhänge nicht ein Artefakt sozialer Erwünschtheit sind.

Zweitens möchten wir die These wagen, dass die Simulation einer Bogus-pipeline-Studie eine effiziente Maßnahme darstellen könnte. Denn nicht nur die Uniformität des Misreporting in Richtung Overreporting, sondern auch das Wording im Test von Belli et al. lässt vermuten, dass beim OR nicht nur Gedächtnisprobleme eine Rolle spielen, sondern vermutete Kontrolleffekte. Zudem wird häufig beobachtet, dass das Overreporting in selbst administrierten und damit schriftlichen Befragungen weniger hoch ist (Tourangeau et al. 295 ff.; Locander et al. 1976: 272). Vor diesem Hintergrund dürften selbst administrierte (schriftliche oder Online-) Befragungen, welche mit der Erhebungsmethode sowie der Frageformulierung gleichzeitig Verbindlichkeit und Kontrollmöglichkeiten suggerieren, wohl am ehesten helfen, das Overreporting zu reduzieren.³¹

Insgesamt erhält man damit den Eindruck, dass auch hier und bei uns kein Weg an mindestens einer größeren und aktuellen Validitätsstudie vorbeiführt, welche Misreporting und Nonresponse gleichzeitig analysiert. Denn nur ein solches Vorgehen erlaubt die Gleichsetzung von weniger Overreporting und mehr Akkuratheit. Zudem bilden Validitätsstudien die Grundlage für jene Maßnahmen, wie sie letzthin auch Jackman (1999) vorgeschlagen und Voogt (2000) angewendet haben: In bestehende Analysen werden Werte aus den Validitätsstudien zum OR als Proxies eingesetzt. Bisher wurde dies allerdings nur mit vereinzelt Validitätsstudien aus Holland,³² aus Australien und den USA getan. Wenn man aber den unterschiedlichen Zugang, die unterschiedlichen Analysemethoden sowie die divergenten Befunde derselben in Betracht zieht, dann ist dieses Vorgehen auch mit Risiken behaftet. Eine effektive Korrektur verlangt präzises Vorwissen, über das wir derzeit noch nicht verfügen. Daher scheint uns auch eine Meta-Analyse der zahlreich vorliegenden Studien ein wichtiger Schritt, und vielleicht inspiriert ja die vorliegende Literaturübersicht zur Durchführung derselben.

31 Dabei ist es wichtig darauf hinzuweisen, dass die Gegenmaßnahmen zum Overreporting sich von jenen gegen ein Underreporting unterscheiden sollten (Locander et al. (1976: 272).

32 Es handelt sich um die holländische Zaanstad-Studie anlässlich der nationalen Wahlen 1998 (Voogt & Saris 2003), die australische Validitätsstudie von 1993 (Vowles et al. 1995) sowie die Validitätsstudien von Traugott für die Jahre 1964-1988 (Traugott 1989).

Literatur

- Abelson, R. P., Loftus E. & Greenwald A. G. (1992). Attempts to Improve the Accuracy of Self-Reports of Voting. In J. M. Tanur (Ed.), *Questions About Questions. Inquiries into the Cognitive Bases of Surveys* (pp. 138-53). New York: Russell Sage Foundation.
- Abramson, P. R. & Claggett, W. (1986). Race-Related Differences in Self-Reported and Validated Turnout in 1984. *Journal of Politics* 48(2), 412-422.
- Adamany, D. & Shelley, M. C. (1980). Encore! The Forgetful Voter. *Public Opinion Quarterly*, 44(2), 234-240.
- Anderson, B. A. & Silver, B. D. (1986). Measurement and Mismeasurement of the Validity of Self-Reported Vote. *American Journal of Political Science*, 30(4), 771-785.
- Andersson, H. E. & Granberg, D. (1997). On the Validity and Reliability of Self-Reported Vote: Validity without Reliability? *Quality and Quantity*, 31, 127-140.
- Atkeson, L. R. (1999). 'Sure, I Voted for the Winner'. Overreport of the Primary Vote for the Party Nominee in the National Election Studies. *Political Behavior*, 21, 197-215.
- Belli, R. F., Traugott, M. W. & Beckmann, M. N. (2001). What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies. *Journal of Official Statistics* 17(4), 479-498.
- Belli, R. F., Traugott, M. W., Young, M., & McGonagle, K. A. (1999). Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. *Public Opinion Quarterly*, 63, 90-108.
- Belli, R. F., Traugott, S. & Rosenstone, S. J. (1994). *Reducing Over-Reporting of Voter Turnout: An Experiment Using a 'Source Monitoring' Framework*. Retrieved 2003, from <http://www.umich.edu/~nes/resources/other/techrpts/belli.htm>
- Bernstein, R. Chadha, A. & Montjoy, R. (2001). Overreporting [Us] Voting: Why It Happens and Why It Matters. *Public Opinion Quarterly*, 65, 22-44.
- Brehm, J. (1993). *The Phantom Respondent. Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.
- Burden, B. C. (2003). Internal and External Effects on the Accuracy of NES Turnout: Reply. *Political Analysis*, 11(2), 193-195.
- Burden, B. C. (2000). Voter Turnout and the National Election Studies. *Political Analysis* 8(4), 389-398.
- Cahalan, D. (1968). Correlates of Respondent Accuracy in the Denver Validity Survey. *Public Opinion Quarterly* 32(4), 607-621.
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and Validity Assessment*. Beverly Hills/London: Sage.
- Cassel, C. A. (2003). Overreporting and Electoral Participation Research. *American Politics Research*, 31, 81-92.

- Clausen, A. R. (1968). Response Validity: Vote Report. *Public Opinion Quarterly*, 32, 588-606.
- Crowne, D. & Marlowe, D. (1964). *The Approval Motive*. New York: Wiley.
- De Leeuw, E. D. & van der Zouwen, J. (1988). Data Quality in Telephone and Face to Face Surveys: A Comparative Meta-Analysis. In P. B. Biemer (Ed.), *Telephone Survey Methodology* (pp. 283-99). New York: Wiley & Sons.
- Eilfort, M. (1994). *Die Nichtwähler. Wahlenthaltung als Form des Wahlverhaltens*. Paderborn: Schöningh.
- Esser, H. (1991). Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „Rational Choice“. In R. Wittenberg (Ed.), *Person – Situation – Institution – Kultur. Günter Büschges Zum 65. Geburtstag* (pp. 59-78). Berlin: Duncker & Humblot.
- Gabler, S. & Häder, S. (Eds.) (2002). *Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland*. Münster: Waxmann.
- Granberg, D. & Holmberg, S. (1992). The Hawthorne Effect in Election Studies: The Impact of Survey Participation on Voting. *British Journal of Political Science*, 22, 240-248.
- Granberg, D. & Holmberg, S. (1991). Self-Reported Turnout and Voter Validation. *American Journal of Political Science*, 35, 448-459.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- Hill, K. Q. & Hurley, P. A. (1984). Nonvoters in Voters' Clothing: The Impact of Voting Behavior Misreporting on Voting Behavior Research. *Social Science Quarterly*, 65(1), 199-206.
- Hyman, H. H. (1945). Do They Tell the Truth? *Public Opinion Quarterly*, 8(4), 557-559.
- Jackman, S. (1999). Correcting Surveys for Non-Response and Measurement Error Using Auxiliary Information. *Electoral Studies*, 18, 7-27.
- Jones, E. E. & Sigall, H. (1971). The Bogus Pipeline: A New Paradigm for Measuring Affect and Attitude. *Psychological Bulletin*, 76, 349-364.
- Kaase, M. (1973). Die Bundestagswahl 1972: Probleme und Analysen. *Politische Vierteljahresschrift*, 14, 145-190.
- Kaase, M. & Bauer-Kaase, P. (1998). Zur Beteiligung an der Bundestagswahl 1994. In M. Kaase & H. D. Klingemann (Eds.), *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 1994* (pp. 85-112). Opladen: Westdeutscher Verlag.
- Karp, J. A. & Brockington, D. (2005). Social Desirability and Response Validity: A Comparative Analysis of Over-Reporting Voter Turnout in Five Countries. *Journal of Politics*, 67(3), 825-840.
- Katosh, J. P. & Traugott, M. (1981). The Consequences of Validated and Self-Reported Voting Measures. *Public Opinion Quarterly*, 45, 519-535.
- Kleinhenz, T. (1995). *Die Nichtwähler. Ursachen der Sinkenden Wahlbeteiligung in Deutschland*. Opladen: Westdeutscher Verlag.
- Krebs, D. (1991). Was ist sozial erwünscht? Der Grad sozialer Erwünschtheit von Einstellungs-items. *ZUMA-Arbeitsbericht 91/18*.

- Lavies, R.-R. (1973). *Nichtwählen als Kategorie des Wahlverhaltens: Empirische Untersuchung zur Wahlenthaltung in historischer, politischer und statistischer Sicht*. Düsseldorf: Droste Verlag.
- Locander, W., Sudman, S. & Bradburn, N. B. (1976). An Investigation of Interview Method, Threat and Response Distortion. *Journal of American Statistical Association*, 71, 269-275.
- Loftus, E. F. et al. (1990). A Tale of Two Questions: Benefits of Asking More Than One Question. *Public Opinion Quarterly*, 54(3), 330-345.
- Martinez, M. D. (2003). Comment on "Voter Turnout and the National Election Studies". *Political Analysis*, 11(2), 187-192.
- McDonald, M. P. (2003). On the Overreport Bias of the National Election Study Turnout Rate. *Political Analysis*, 11, 180-186.
- Miller, M. (1952). The Waukegan Study of Voter Turnout Prediction. *Public Opinion Quarterly*, 16, 381-398.
- Mummendey, H. D. (1999). *Die Fragebogen-Methode. Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung*. 2. Ausgabe. Göttingen/Bern/Toronto/Seattle: Hogrefe.
- Parry, H. J. & Crossley, H. M. (1950). Validity of Responses to Survey Questions. *Public Opinion Quarterly*, 12, 585-598.
- Phillips, D. L. (1973). *Abandoning Method*. San Francisco: Jossey-Bass.
- Presser, S. (1984). Is Inaccuracy on Factual Survey Items Item-Specific or Respondent-Specific? *Public Opinion Quarterly*, 48, 344-355.
- Presser, S. (1990). Can Changes in Context Reduce Vote Overreporting in Surveys? *Public Opinion Quarterly*, 54, 586-593.
- Presser, S. & Traugott, M. (1992). Little White Lies and Social Science Models: Correlated Response Error in a Panel Study of Voting. *Public Opinion Quarterly*, 54, 77-86.
- Presser, S., Traugott, M. W. & Traugott, S. M. (1990). *Vote 'over' Reporting in Surveys: The Records or the Respondents?* National Election Study Technical Report 010157. Michigan.
- Rattinger, H. & Wagner, S. (2003). *Germany: Supplementing Sense of Duty by Cognitive Mobilisation*. Bamberg: University of Bamberg.
- Renzetti, C. M. & Lee, R. (Eds.) (1993). *Researching Sensitive Topics*. Newbury Park: Sage.
- Richman, W. L., Kiesler, S., Weisband, S., Drasgow, F. (1999). A Meta-Analytical Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews. *Journal of Applied Social Psychology*, 84(5), 754-775.
- Riklin, A. & Kley, R. (1981). *Stimmabstinenz und direkte Demokratie. Ursachen – Bewertungen – Konsequenzen*. Bern/Stuttgart: Paul Haupt.
- Rogers, T. F. (1976). Interviews by Telephone and in Person: Quality of Responses and Field Performance. *Public Opinion Quarterly*, 40, 51-65.

- Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen*. Opladen: Leske + Budrich.
- Scholl, A. (1993). *Die Befragung als Kommunikationssituation. Zur Reaktivität im Forschungsinterview*. Opladen: Westdeutscher Verlag.
- Schwarz, N. & Scheuring, B. (1992). Selbstberichtete Verhaltens- und Symptommhäufigkeiten: Was Befragte aus Antwortvorgaben des Fragebogens lernen. *Zeitschrift für Klinische Psychologie*, 21, 197-208.
- Sigelman, L. (1982). The Nonvoting Voter in Voting Research. *American Journal of Political Science*, 26(1), 47-56.
- Silver, B. D. et al. (1986). Who Overreports Voting? *American Political Science Review* 88(2), 613-624.
- Steiner, J. (1965). Überprüfung von Interview-Ergebnissen über die Stimm- und Wahlbeteiligung durch amtliche Angaben. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 17, 234-244.
- Stöber, J. (1999). Die Soziale-Erwünschtheits-Skala-17 (Ses-17): Entwicklung und erste Befunde zur Reliabilität und Validität. *Diagnostica*, 45(4), 173-177.
- Sudman, S. & Bradburn, N. B. (1974). *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.
- Swaddle, K. & Heath, A. (1989). Official and Reported Turnout in the British General Election of 1987. *British Journal of Political Science*, 19, 527-541.
- Tourangeau, R., Rips, L. J. & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Traugott, M. (1989). *Validating Self-Reported Vote 1964-1988. Technical Report No. 34 National Election Studies*. Michigan: ISR University of Michigan.
- Traugott, M. W. & Katosh, J. P. (1979). Response Validity in Survey of Voting Behavior. *Public Opinion Quarterly*, 43, 359-377.
- Traugott, S. (1989). *Validating Self-Reported Vote: 1984 - 1988, National Election Study*. Technical Report 010152. Michigan.
- Van der Kolk, H. & Aarts, K. (2001). *Turnout Bias. Non-Response and Stimulus Effects in the Dutch Parliamentary Election Study 1998*. Paper presented at the Annual Meeting of the American Political Science Association, San Francisco.
- Verband Schweizer Marketing- und Sozialforscher (SMS) (2001). *Jahrbuch 2001*. Berneck: SMS.
- Volgy, T. J. & Schwarz, J. E. (1984). Misreporting and Vicarious Political Participation at the Local Level. *Public Opinion Quarterly*, 48, 757-765.
- Voogt, R. (2000). *An Alternative Approach to Correcting Response and Nonresponse Bias in Election Research*. Amsterdam: University of Amsterdam.

- Voogt, R. & Saris, W. E. (2003). To Participate or Not to Participate: The Link between Survey Participation, Electoral Participation, and Political Interest. *Political Analysis*, 11(2), 164-179.
- Vowles, J. et al. (1995). *Towards Consensus: The 1993 Election in New Zealand and the Transition to Proportional Representation*. Auckland: Auckland University Press.
- Wolfinger, R. & Rosenstone, S. J. (1980). *Who Votes?* New Haven: Yale University Press.
- Wright, G. C. (1993). Errors in Measuring Vote Choice in the National Election Studies, 1952-1988. *American Journal of Political Science*, 37, 291-316.

Abkürzungsverzeichnis

AHV	Alters- und Hinterlassenenversicherung
ANES	American National Election Study
GfS	Gesellschaft für praktische Sozialforschung
IHA-GfK	Meinungsforschungsinstitut Schweiz, Hergiswil, Marktleader
IPSO	Meinungsforschungsinstitut Schweiz, Zürich; gehört mittlerweile zu IHA
IV	Invalidentversicherung
JSTOR	The Scholarly Journal Archive
OR	Overreporting
SIDOS	Schweizerischer Informations- und Daten-Archivdienst für die Sozialwissenschaften
SNF	Schweizerischer Nationalfonds
VOX/VOX-Analyse	Nachbefragung zu den eidgenössischen Urnengängen, 1977 ff.

Korrespondenzadressen

Prof. Dr. Sibylle Hardmeier
Forschungsprofessorin
Wissenschaftszentrum Berlin für
Sozialforschung (WZB)
Reichpietschufer 50
10785 Berlin
E-mail: hardmeier@wz-berlin.de

Marie-Christine Fontana, lic. phil.
Institut d'Etudes Politiques et Internationales
Bât. Humense
Université de Lausanne
1015 Lausanne
Schweiz
E-mail: Marie-Christine.Fontana@unil.ch

ZUR WIRKUNG UNTERSCHIEDLICHER MATERIELLER INCENTIVES IN POSTALISCHEN BEFRAGUNGEN EIN LITERATURBERICHT

EFFECTS OF INCENTIVES ON MAIL SURVEY RESPONSES A REVIEW OF RESEARCH LITERATURE

FRED BERGER

Materielle Anreize, so genannte Incentives, haben sich neben dem Versand von Erinnerungsschreiben in vielen Methodenstudien als die wirksamste Maßnahme zur Steigerung der Ausschöpfungsquote in postalischen Befragungen erwiesen.

Im Beitrag werden die gängigsten theoretischen Konzepte zur Erklärung der Wirkungsweise von Incentives dargestellt und ein Überblick über den Stand der Methodenforschung in diesem Bereich gegeben. Der Beitrag mündet in einigen Empfehlungen zur Verwendung von materiellen Anreizen in der Forschungspraxis.

A large number of studies prove prepaid monetary and non-monetary incentives as well as follow-up contacts to be the most effective procedures to increase mail survey response rates.

This paper describes established theories concerning the effectiveness of incentives and reviews research results in this area. It concludes with a few recommendations on the practical use of incentives.

1 Einleitung

Schriftliche und postalische Befragungen besitzen in der sozialwissenschaftlichen Forschung als Erhebungsverfahren einen relativ hohen Verbreitungsgrad. Bei den in der Datenbank FORIS¹ in den letzten drei Jahren eingetragenen deutschsprachigen Forschungsprojekten beträgt das Verhältnis von schriftlichen und postalischen Befragungen

1 Forschungsinformationssystem Sozialwissenschaften (FORIS) des Informationszentrums Sozialwissenschaften (IZ) in Bonn

gegenüber mündlich oder telefonisch durchgeführten Einzelinterviews beispielsweise 543:1147. Trotz dieser großen Verbreitung führt insbesondere die postalische Befragung eher ein „Schattendasein“ (vgl. Thoma & Zimmermann 1996: 142). In der Forschungspraxis wird ihr Einsatz häufig als Kompromiss gesehen, der vorwiegend auf Grund des finanziell Machbaren eingegangen werden muss und weniger dem methodisch Wünschbaren entspricht (vgl. Blasius & Reuband 1996: 35). Face-to-Face- und telefonische Befragungen gelten bei kompetenter Durchführung gemeinhin als geeigneter, um die Befragten zur Teilnahme an der Untersuchung zu motivieren, ihnen Hilfestellungen bei Verständnisproblemen zu geben und um auf der Basis einer hohen Ausschöpfungsquote eine repräsentative Bevölkerungsstichprobe zu erreichen (Groves et al. 2004). Postalischen Befragungen wird demgegenüber eine generell geringere Ausschöpfungsquote und eine verminderte Teilnahmebereitschaft von älteren Menschen (Kaldenberg, Koenig & Becker 1994) sowie von lese- und schreibungeübten Personen zugeschrieben (Dillman 1978).

Die angeführten Bedenken gegenüber postalischen Befragungen erweisen sich angesichts des aktuellen Methodenwissens jedoch nur mehr teilweise als begründet (vgl. Blasius & Reuband 1996: 35). Die Forschung hat in den letzten Jahrzehnten verschiedene Möglichkeiten zur Steigerung der Teilnahmebereitschaft entwickelt und optimiert und dadurch ein beträchtliches Potential zur Behebung früherer Defizite geschaffen (zusammenfassend Klein & Porst 2000). Am effektivsten haben sich der Versand von Erinnerungsschreiben und die Vergabe materieller Anreize (so genannte Incentives) erwiesen (Furse, Stewart & Rados 1981; Yu & Cooper 1983; James & Bolstein 1990; Yammarino, Skinner & Childers 1991; Church 1993; Dodd 1998; Porst, Ranft & Ruoff 1998). Durch den gezielten Einsatz mehrerer Erinnerungsschreiben können die sonst typischerweise zwischen 30 und 50 Prozent liegenden Ausschöpfungsquoten (vgl. Hopkins & Stanley 1981, zitiert in Klein & Porst 2000: 15) beispielsweise auf ein Niveau von 60 bis 75 Prozent angehoben werden (Dillman 1983: 360; Hippler & Seidel 1985: 45; Thoma & Zimmermann 1996: 150; Dillman 2000; Diaz de rada 2005). Mit dem Einsatz von Incentives kann eine weitere, von der Beschaffenheit des Anreizes und dem Anforderungsgehalt der Studie für die Befragten abhängige Verbesserung der Teilnahmebereitschaft erzielt werden (siehe Pkt. 3). Es liegen damit Teilnahmequoten im Bereich des Möglichen, die mit Face-to-Face-Befragungen vergleichbar sind (Porst 1996; Reuband & Blasius 1996) und in einzelnen Fällen sogar in die Nähe von Telefonstudien kommen.

2 Theoretische Erklärungen zur Wirkung von Incentives

2.1 Die Bedeutung von Incentives im Rahmen einer allgemeinen Theorie des Teilnahmeverhaltens

Wenn in neueren Arbeiten versucht wird, die Wirkung materieller Anreize zu erklären, so geschieht dies häufig im Rahmen einer allgemeinen Theorie des Teilnahmeverhaltens, die entweder als Rational-Choice-Theorie formuliert ist oder sich als solche rekonstruieren lässt. Der Entscheid für oder gegen die Teilnahme an einer Befragung wird in diesen allgemeinen Theorien als Folge einer Kosten-Nutzen-Abwägung des Individuums auf der Grundlage aller subjektiv bedeutsamen Anforderungen und Gratifikationen einer Untersuchung verstanden (Esser 1986a: 321; Goyder 1987; Schnell 1997: 158ff). Eine Teilnahme ist demzufolge dann zu erwarten, wenn der individuelle Gewinn die durch die Beteiligung entstehenden Kosten übersteigt bzw. wenn der Austausch aus der Perspektive des Probanden als gerecht wahrgenommen wird.

Von einem persönlichen Nutzen wird dann ausgegangen, wenn mit der Teilnahme an postalischen Befragungen z.B. eine Steigerung des Selbstwerts (weil man wichtig ist und an etwas Wichtigem teilnimmt), die Möglichkeit zur Befriedigung altruistischer Motive, Lernerfahrungen (mit interessanten Themen konfrontiert werden) oder eben auch materielle Anreize verbunden sind.²

Zu den Kosten werden vorwiegend die Belastungen gezählt, die sich aus dem Befragungsvorgang selbst ergeben (z.B. Zeitverlust, Offenbarung unvorteilhafter Informationen über die eigene Person) oder die aus Befürchtungen über die missbräuchliche Verwendung der Daten entstehen (vgl. Schnell 1997: 166).

2.2 Incentives als Vertrauensbeweis und symbolischer Akt der Aufwandsentschädigung

Die erste explizite Kosten-Nutzen-Theorie des Teilnahmeverhaltens in postalischen Befragungen wurde von Dillman (1978) entworfen, nachdem lange Zeit individuelle Eigenschaften und Wertvorstellungen der Probanden für die Teilnahme bzw. Verweigerung verantwortlich gemacht wurden. In seiner auf den Arbeiten von Homans (1961), Blau (1964) und Thibaut & Kelley (1959) beruhenden sozialen Austauschtheorie misst Dillman

2 Esser (1986b: 39) gibt zu bedenken, dass bei Befragungen in vielen Fällen sowohl die Nutzen- wie auch die Kostenaspekte nur gering ausgeprägt sein dürften, so dass die Entscheidung zur Teilnahme auch aus einer Art Indifferenz heraus erfolgen kann. Wie in solchen Situationen üblich, spielen dann auch relativ periphere und zufällige Faktoren (wie z.B. die Stimmung) eine Rolle.

im Hinblick auf die Erzeugung von Teilnahmebereitschaft neben der Minimierung der Kosten und der Maximierung des Nutzens besonders dem Aufbau von Vertrauen bei den Befragten eine große Bedeutung bei. Da in postalischen Befragungen für die Probanden meist keine Möglichkeit besteht, die von den Untersuchern gemachten Versprechen (wie z.B. die Zustellung eines Ergebnisberichts, die sinnvolle Verwendung der Ergebnisse und die vertrauliche Behandlung der Daten) von vornherein sicherzustellen, stellt das Vertrauen in eine gerechte Austauschbeziehung eine unabdingbare Voraussetzung für eine Kooperation dar (Dillman 1978: 12-16). Incentives werden auf diesem Hintergrund primär als Zeichen des Vertrauens und der Wertschätzung gesehen und kommen erst in zweiter Linie wegen ihres materiellen Nutzens für die Probanden zum Einsatz (Dillman 1978: 16). Sie sollen eine soziale Austauschbeziehung einleiten und bei den Befragten die Verpflichtung generieren, dem Wunsch der Untersucher zu entsprechen. Die Teilnahme an der Studie stiftet durch die Erfüllung der Erwartung den Probanden dann einen zusätzlichen Nutzen (vgl. Arzheimer & Klein 1998: 8)³.

Der von Dillman postulierte Verpflichtungscharakter von Incentives wird heute vor allem auf dem Hintergrund der Theorie reziproken Handelns diskutiert (Gouldner 1960). In ihr wird von einer kulturübergreifend gültigen Norm der Reziprozität ausgegangen, die in ihrer universalen Verbreitung zwei minimale Bestandteile enthält: Man soll zum einen jenen helfen, von denen man selbst Hilfe erhalten hat und man soll zum anderen Personen, die einem geholfen haben, kein Leid antun. Die Schlussfolgerung ist klar: Wenn man will, dass einem geholfen wird, muss man auch anderen helfen. In dieser Ausprägung gehört die Norm der Reziprozität zu den Grundkomponenten jedes gesellschaftlichen Moralkodes und trägt wesentlich zur Stabilisierung sozialer Systeme bei (Gouldner 1960: 171, 173). Sie ist als ein internalisierter Wert, als ein Produkt der Sozialisation zu verstehen, verbindet aber gleichzeitig auch egoistische Motive mit dem Dienst an anderen (Gouldner 1960: 174).

3 Mit der austauschtheoretischen Konzeption bzw. den Rational-Choice-Theorien ist ein Rahmen gegeben, in welchen sich im Prinzip auch weitere im Zusammenhang mit dem Teilnahmeverhalten in Befragungen und der Wirkungsweise von Incentives relevante, stärker psychologisch orientierte oder die Bedeutung von sozialen Normen betonende Erklärungsansätze einordnen lassen (vgl. z.B. die Wirkmechanismen in den Konzeptionen von Groves, Cialdini & Couper 1992; Cavusgil & Elvey-Kirk 1998). Die im Folgenden in Erweiterung zur Dillmanschen Konzeption dargestellten Ansätze können dementsprechend als habitualisierte Verhaltenstendenzen oder als subjektiv rationale Handlungsheuristiken (die durch materielle Anreize ausgelöst werden) unter einer allgemeinen handlungstheoretischen Perspektive des Teilnahmeverhaltens subsumiert werden (vgl. z.B. Schnell 1997). Aus psychologischer Perspektive wäre eine Einordnung unter dem Dach der kognitiven Dissonanztheorie (vgl. Festinger 1957) aber genauso plausibel. Entsprechende Umsetzungen finden sich z.B. bei Gelb (1975), Furse & Stewart (1982) und Cavusgil & Elvey-Kirk (1998).

Mit der Annahme einer universalen Gültigkeit ist allerdings nicht die Meinung verbunden, reziprokes Handeln erfolge ohne jegliche Voraussetzung (Gouldner 1960: 171). Die Verpflichtung zur Reziprozität ist immer in Abhängigkeit vom subjektiven Wert der erhaltenen Leistung zu sehen. Es können zudem auch bei gleichem subjektivem Wert einer Dienstleistung unterschiedliche Vorstellungen über die Verbindlichkeit reziproken Handelns existieren.⁴

Auf dem Hintergrund der Reziprozitätsregel und den Überlegungen Dillmans (1978) ist zu erwarten, dass materielle Anreize dann die größte Wirkung entfalten, wenn sie, ohne an Bedingungen geknüpft zu sein, bereits mit dem ersten Anschreiben verschickt werden. Die Wirkungsweise von auf diese Art unaufgefordert gegebenen Geschenken besteht nach Meinung verschiedener Autoren (z.B. Gelb 1975; Furse et al. 1981; Furse & Stewart 1982: 379-378) darin, dass sie bei den Probanden kognitive Spannungszustände und Zwangslagen im Sinne der Dissonanztheorie von Festinger (1957) erzeugen, die dann durch eine Beteiligung an der Studie wieder aufgelöst werden können. Das innere Dilemma ergibt sich für die Probanden daraus, dass das Incentive auf Grund der verinnerlichten Norm zur Gegenseitigkeit und des mit ihm dokumentierten Vertrauens nicht ohne eine Gegenleistung behalten werden kann, sofern es einen gewissen materiellen oder subjektiven Wert besitzt. Es kann auch nicht einfach ignoriert und weggeworfen werden, da dies in den meisten Fällen ebenfalls keine mit den eigenen Wertvorstellungen zu vereinbarende Alternative darstellt. Die Teilnahme an der Studie bildet so häufig die beste Möglichkeit, um die kognitive Spannung zu beseitigen und in Übereinstimmung mit internalisierten Werten zu handeln.

Dillman (1978: 16) meint, dass es für die angeschriebenen Probanden umso einfacher werde, die Beteiligung an einer Untersuchung abzulehnen, je näher der Wert des Incentives der verlangten Arbeit komme und je mehr die Austauschhandlung den Charakter eines ökonomischen Austauschs annehme. Durch ein unangemessen hohes Incentive kann die Austauschbeziehung in den Augen der Probanden ihre vermeintliche Freiwilligkeit verlieren und als rein ökonomische Handlung Reaktanz provozieren oder einen einfachen kognitiven Ausstieg ermöglichen (vgl. Arzheimer & Klein 1998: 8; Martin, Abreu & Winters 2001: 282). Kleine symbolische Geschenke oder Geldbeträge bewähren sich nach Dillman am besten und ermöglichen in der Regel eine beträchtliche Rücklaufsteigerung.

4 Muir & Weinstein (1962) konnten diesen Umstand in einer Untersuchung mit Unterschicht- und Mittelschichtangehörigen in den Vereinigten Staaten nachweisen. Das Denken und Handeln der Mittelschicht erwies sich in ihrer Studie in stärkerem Maße durch Erwartungen an Gegenseitigkeit geleitet als jenes der Unterschicht. Die Autoren fassen ihre Ergebnisse in der Formel zusammen, dass Personen der Mittelschicht sich bei erhaltener Leistung zur Erwidern verpflichtet, Personen der Unterschicht hingegen eher dankbar fühlen (S. 538).

2.3 Incentives als ökonomisch bedeutsame Entlohnung

In Erweiterung bzw. Abgrenzung zu dieser Position der primär symbolischen Aufwand-entschädigung wird von verschiedenen Autoren der Bezahlungsaspekt von Incentives als Motiv für die Teilnahme an einer Befragung stärker in Betracht gezogen (vgl. Biener & Kidd 1994; Shettle & Mooney 1999). Incentives vermögen danach auch (oder vor allem) wegen ihres materiellen Wertes die Kosten-Nutzen-Kalkulation der Probanden zu Gunsten einer Partizipation zu beeinflussen. Sie haben in diesen Überlegungen die Funktion einer teilweisen oder vollständigen materiellen Entschädigung für die geleistete Arbeit. Von Relevanz sind sie vor allem dann, wenn in einer Untersuchung ein besonderer Aufwand verlangt wird oder wenn den Probanden auf andere Weise kein ausreichender persönlicher Nutzen einer Teilnahme entsteht.

Geht man dieser Logik entsprechend bei den Probanden von einem streng rationalen Nutzenkalkül aus, dann müssten Incentives unter mehr oder weniger expliziter Deklaration des Entschädigungsaspekts (z.B. als versprochene Bezahlung) mit zunehmendem materiellem Wert an Wirkung gewinnen (vgl. Warriner, Goyder, Gjertsen, Hohner & Mc Spurren 1996; Diekmann & Jann 2001). Es ist aber nicht zu erwarten, dass durch eine ausreichend hohe Bezahlung tatsächlich eine nahezu vollständige Ausschöpfung erreicht werden kann. Es scheint vielmehr wahrscheinlich, dass ähnliche Normen und psychologische Prozesse wie oben beschrieben regulierend wirksam werden und eine Rücklaufsteigerung nur bis zu einem Grenzwert hin ermöglichen. Diese Annahme wird beispielsweise durch Überlegungen zur Verteilungsgerechtigkeit (equity theory: Homans 1961; Adams 1965) und die Theorie psychologischer Reaktanz gestützt (Brehm 1966; Brehm & Cole 1966). Zu große Geschenke oder Geldbeträge können danach als ungerechtfertigt und unverhältnismäßig angesichts des verlangten Aufwands und der anderweitig vorhandenen Nutzenaspekte beurteilt oder als ungebührlicher Versuch der Einengung der persönlichen Entscheidungsfreiheit verstanden werden („Bestechung“). Die Teilnahmeverweigerung ist dann als Konsequenz des Versuchs zu sehen, die eigenen Freiheitsspielräume wieder herzustellen oder in Übereinstimmung mit subjektiv bedeutsamen Wertvorstellungen zu handeln. Große Incentives können bei den Probanden außerdem den Eindruck erwecken, dass die Untersucher ihre Studie selbst als zu wenig interessant und intrinsisch motivierend einschätzen, um ohne eine deutliche materielle Entschädigung eine ansprechende Beteiligungsbereitschaft zu erzeugen (vgl. Lynn 2001).

3 Befunde aus der Methodenforschung

3.1 Monetäre und non-monetäre Anreize von symbolischer Bedeutung

In welcher Weise Incentives wirken, ob als rein symbolische Aufwandentschädigung oder im Sinne einer ökonomisch bedeutsamen Entlohnung (oder allenfalls durch beide Wirkmechanismen gleichzeitig), lässt sich nicht alleine auf Grund von theoretischen Überlegungen entscheiden. Es bedarf dazu empirischer Unterstützung. Die mittlerweile umfangreiche Forschungsliteratur präsentiert sich jedoch nicht frei von Widersprüchen und offenen Fragen.

In vielen Belangen lassen sich dennoch auf dem Hintergrund einer großen Zahl von Studien verlässliche Aussagen machen. So gibt es z.B. kaum gegensätzliche Ergebnisse und Meinungen zur positiven Wirkung von finanziellen Anreizen und von non-monetären Geschenken – vorausgesetzt, sie werden beim Erstkontakt, und ohne dass sie an Bedingungen geknüpft sind, vergeben. Es kann generell davon ausgegangen werden, dass sie den Rücklauf in postalischen und in mündlichen Befragungen in bedeutsamer Weise zu steigern vermögen (z.B. Chebat & Cohen 1993; Church 1993; Martinez-Ebers 1997; Arzheimer & Klein 1998; Dodd 1998; Porst et al. 1998; Singer, Groves & Corning 1999; Diekmann & Jann 2001; McConaghy & Beerten 2003; Simmons & Wilmot 2004). Sie bewähren sich sowohl in Studien mit geringem als auch mit hohem Anforderungsgehalt für die Probanden. Ihre Wirkung steigt jedoch mit dem Aufwand, der für die Befragten durch die Teilnahme entsteht (Singer 1998: 10-11; Singer, van Hoewyk, Gebler, Raghunathan & McGonagle 1999: 224) und mit dem Anteil an Probanden, die der Untersuchung anfänglich uninteressiert oder ablehnend gegenüber stehen (Groves, Singer & Corning 2000; Trussell & Lavrakas 2004: 361).

Versprochene Incentives haben sich demgegenüber als weniger wirkungsvoll erwiesen. Church (1993: 75) kommt nach einer (Meta-)Analyse von 19 Teilnahmequoten sogar zum Schluss, dass (im Falle der Teilnahme an der Studie) in Aussicht gestellte non-monetäre und finanzielle Entschädigungen den mit ihnen verbundenen Aufwand nicht rechtfertigen. Sie bewirken in vielen Untersuchungen keine bedeutsame Rücklaufsteigerung gegenüber der Nullvariante (James & Bolstein 1992; Church 1993; Diekmann & Jann 2001) oder schneiden deutlich schlechter ab als ein vom Wert her ungefähr vergleichbares mit dem Erstkontakt und ohne Bedingungen gegebenes (prepaid) Incentive (Gelb 1975; Yu & Cooper 1983; Berry & Kanouse 1987). Auch alternative Formen von in Aussicht gestellten Belohnungen wie zum Beispiel die Teilnahme an einer Lotterie oder die Zusage einer Spende an eine gemeinnützige Organisation (altruistisches Incentive) ermöglichen in sozialwissenschaftlichen Untersuchungen in der Regel nicht den selben Ertrag (bzw. die

selbe Effizienz in der Nutzung der vorhandenen Mittel) wie vorausbezahlte Incentives (Furse & Stewart 1982; Hubbard & Little 1988; Warriner et al. 1996)⁵.

In den Vereinigten Staaten und in Kanada, wo bisher weit mehr Forschung als im europäischen Raum zur Frage von Incentives betrieben wurde, haben sich im Rahmen von postalischen Befragungen besonders mit dem Erstanschreiben vergebene monetäre Anreize als wirksam und robust erwiesen. Dabei haben sich (im Sinne der Reziprozitätsnorm) bereits kleine, eher symbolische Geldbeträge gut bewährt (Furse & Stewart 1982; Yu & Cooper 1983; James & Bolstein 1990, 1992; Church 1993; Warriner et al. 1996: 546; Trussell & Lavrakas 2004). Die rücklaufsteigernde Wirkung von finanziellen Anreizen übertrifft in diesen Breitengraden jene von non-monetären Geschenken (wie z.B. Kugelschreiber, Anstecknadeln, Briefmarken) meist deutlich (Hansen 1980; Church 1993; Simmons & Wilmot 2004). Church (1993: 67-71) errechnete in der bereits erwähnten Metaanalyse von postalischen Befragungen (nun auf der Basis von 43 resp. 12 Teilnahmequoten) eine mittlere Rücklaufsteigerung von 19.1 Prozent für prepaid monetäre Anreize (mit einem durchschnittlichen Kaufwert von zirka \$ 1.40 im Jahre 1989) und von 7.9 Prozent für kleine dem Erstversand beigelegte Sachgeschenke⁶. Die grundsätzliche Überlegenheit von im Voraus gegebenen monetären gegenüber nicht-monetären Anreizen konnte von Singer et al. (1998; 1999; 2000)⁷ auch für Telefon- und Face-to-Face-Studien (unter Kontrolle des materiellen Wertes) belegt werden. Sie könnte u.a. darin begründet sein, dass finanzielle Anreize im allgemeinen einen größeren Verpflichtungsgrad (und eine größere kognitive Dissonanz: vgl. Furse & Stewart 1982: 379) erzeugen, weil sie (bis zu einem gewissen Betrag hin) sowohl als Geschenk als auch als Versuch zur Aufwandschädigung verstanden werden können, während non-monetäre Incentives „nur“ den Charakter eines Geschenkes aufweisen (vgl. Martinez-Ebers 1997). Non-monetäre Anreize können allenfalls bei einer homogenen

5 Gleiches gilt für materielle Anreize, die zur Umstimmung von Verweigerern zu einem späteren Zeitpunkt im Untersuchungsverlauf eingesetzt werden (refusal conversion: vgl. Groves, Singer, Corning & Bowers 1999; Singer, Groves et al. 1999).

6 Yu & Cooper (1983: 40) haben in einer anderen Metaanalyse unter Vergleich von 55 Rücklaufquoten eine durchschnittliche Steigerung von 15.6 Prozent für beim Erstkontakt überreichte finanzielle Anreize errechnet. 92 Prozent der in dieser Metaanalyse einbezogenen (prepaid und promised) monetären Incentives bewegten sich im Bereich von \$ 0.10 bis \$ 5.00 (eigene Berechnungen auf Grund von Angaben in Yu & Cooper 1983: 40, Tab. 3). Die von den Autoren analysierten Studien wurden im Zeitraum von 1967 bis 1981 publiziert. Bei beinahe 80 Prozent handelte es sich um postalische Befragungen.

7 Die Metaanalyse von Singer et al. (1998; 1999) belegt auch die Überlegenheit von prepaid gegenüber promised Incentives für telefonische und mündliche Befragungen. Sie umfasst 39 amerikanische und kanadische Untersuchungen, die bis Ende der 90er Jahre des letzten Jahrhunderts durchgeführt wurden. 69 Prozent davon waren mündliche Befragungen, der Rest fand über Telefon statt.

Stichprobe von Vorteil sein, wenn sie für die Probanden erwiesenermaßen von besonderem subjektivem oder materiellem Wert sind. In heterogenen Stichproben besteht jedoch immer die Gefahr, dass sie auf Grund ihrer Beschaffenheit für viele Probanden nicht sehr attraktiv oder brauchbar sind⁸.

Im deutschsprachigen Raum wurden im akademischen Bereich bisher vor allem Untersuchungen zur Wirkung von kleineren Geschenken in Form von Briefmarken (im Wert von DM 4 bzw. vier Standardbriefen: Harkness, Mohler, Schneid & Christoph 1998; Porst 1999) oder von Telefonkarten (in der Höhe von DM 6 bzw. CHF 10: Arzheimer & Klein 1998; Diekmann & Jann 2001) durchgeführt. Die dem Erstkontakt beigelegten geldnahen Sachgeschenke bewirkten dabei eine Steigerung der Ausschöpfungsquote von 5 bis 12.3 Prozentpunkten. Ein Vergleich von prepaid und promised Incentives (in der Form von Telefonkarten) durch Diekmann & Jann (2001) ergab zudem klare Vorteile für die Prepaid-Variante. Die Ergebnisse deuten darauf hin, dass im deutschen Sprachraum grundsätzlich von vergleichbaren Effekten und Wirkmechanismen wie im amerikanischen Bereich ausgegangen werden kann, dass nämlich bereits materielle Anreize von geringem bis moderatem Wert mit mehr symbolischer als ökonomischer Bedeutung zu einer nicht zu vernachlässigenden Rücklaufsteigerung führen, wenn sie gleich zu Beginn der Untersuchung gegeben werden.

Dieser Befund bedeutet nun allerdings nicht, dass der finanzielle Wert eines Incentives ganz ohne Belang für die Teilnahme an einer Untersuchung ist. Verschiedene (vor allem amerikanische) Studien legen es im Gegenteil nahe, dass ökonomische Überlegungen bei den Probanden ebenfalls eine Rolle spielen.

3.2 Möglichkeiten der Rücklaufsteigerung und Grenzen der Wirksamkeit in Abhängigkeit vom finanziellen Wert eines Incentives

In der angelsächsischen Forschungsliteratur besteht heute allgemein der Konsens, dass die Rücklaufquote in einer Untersuchung mit zunehmender Höhe eines Incentives steigt (Church 1993; Dodd 1998; Singer, van Hoewyk et al. 1999)⁹. Obwohl in der Regel bereits relativ kleine prepaid Incentives eine deutliche Verbesserung der Teilnahmebereitschaft gegenüber einer Nullvariante bewirken, lässt sich demzufolge durch eine Erhöhung des materiellen Wertes in vielen Fällen der Rücklauf optimieren. Umstritten ist allerdings, ob der Zuwachs in linearer Abhängigkeit vom finanziellen Wert erfolgt oder in abnehmender

8 Dies dürfte grundsätzlich auch für geldnahe Incentives wie z.B. Briefmarken und Telefonkarten zutreffen.

9 Für unseren kulturellen Kontext fehlen bis anhin systematische Studien zur Wirkung von Incentives mit unterschiedlichem materiellem Wert.

Rate ansteigt und nach einer gewissen Höhe zum Stillstand kommt (model of diminishing returns). Während ein strikt lineares Modell als Hinweis für rationale Nutzenmaximierungsabsichten auf Seiten der Probanden gedeutet werden kann, dürfte das Grenzwertmodell eher den Versuch der Untersuchungsteilnehmer abbilden, ihren subjektiven Nutzen ins Gleichgewicht mit den Anforderungen bzw. persönlichen Kosten der Studie zu bringen. Beide Modelle schließen nicht aus, dass für einen beträchtlichen Teil der Probanden unabhängig vom ökonomischen Wert primär die mit dem Geschenk zum Ausdruck gebrachte Wertschätzung als Motiv für die Teilnahme an der Studie handlungsleitend war.

In der Literatur finden sich für beide Modelle empirische Belege. Yu & Cooper (1983) sowie Singer et al. (1998; 1999) stützen mit ihren großen Metaanalysen von postalischen, telefonischen und mündlichen Befragungen die erste Position. Sie ermittelten einen linearen Zuwachs der Rücklaufquote bis hin zum (damaligen) Kaufwert von \$ 50 bzw. \$ 100.¹⁰ Church (1993) hat in seiner Metaanalyse ebenfalls einen linearen Anstieg in Abhängigkeit vom Wert des Incentives errechnet. In seine Berechnungen sind jedoch nur kleine bis moderate Geldbeträge eingeflossen (bis zum Kaufwert von \$ 9.29).¹¹ Ein monotoner Zuwachs konnte auch in der Studie von Trussel & Lavrakas (2004) festgestellt werden. In ihr wurden kleinere prepaid Incentives in der Höhe von 1 bis 10 Dollar in inkrementellen Schritten von einem Dollar eingesetzt.

Die These eines nicht-linearen Anstiegs mit abnehmender Wirksamkeit wird demgegenüber u.a. durch die Untersuchungen von Armstrong (1975), Fox et al. (1988), James & Bolstein (1992), Warriner et al. (1996) und Martin et al. (2001) getragen. In den drei letztgenannten Studien ergaben sich auch deutliche Hinweise auf die Existenz von Grenznutzen-Schwellen im oberen Bereich. Bei Warriner (1996) lag diese Schwelle bereits bei \$ 5, während sie in den anderen beiden Untersuchungen erst mit \$ 20 erreicht war. Eine Erhöhung des finanziellen Anreizes von \$ 20 auf \$ 40 bewirkte in der Untersuchung von James & Bolstein (1992) sogar eine signifikante Abnahme der Teilnahmebereitschaft auf das Niveau von prepaid Incentives mit geringem bis moderatem Wert. Die Autoren erklä-

10 Der Zusammenhang zwischen der Höhe der (prepaid und promised) monetären Incentives und den gewichteten Rücklaufquoten beträgt in der Studie von Yu & Cooper (1983: 39-40) 0.61. Singer et al. (1998; 1999) errechneten in ihrer Metaanalyse eine durchschnittliche Rücklaufsteigerung von 0.33 Prozent für jeden zusätzlichen Dollar an Incentive. Die Berechnungen von Singer et al. sind standardisiert auf den Kaufwert von 1983. Zu den Analysen von Yu und Cooper liegen keine entsprechenden Angaben vor. In beiden Metaanalysen ist der Anteil an Studien mit kleinen oder moderaten Incentives wesentlich größer als der von Untersuchungen, die große Geldbeträge verwendeten.

11 Der Betrag ist standardisiert auf den Kaufwert von 1989. Die Korrelation zwischen der Höhe der prepaid monetären Incentives und den gewichteten Rücklaufquoten bzw. Effektstärken beträgt in dieser Metaanalyse 0.45.

ren sich diesen Grenz- bzw. Umkehreffekt durch den Umstand, dass der Betrag von vielen Probanden als zu hoch und unangemessen eingestuft wurde und entsprechend Ablehnung provozierte (James & Bolstein 1992; Martin et al. 2001).

Neben einem oberen Grenzwert existiert bei der Vergabe von materiellen Anreizen offenbar auch eine untere Grenze der Wirksamkeit. Sehr kleine bzw. subjektiv bedeutungslose Incentives können in Verbindung mit einer für die Probanden aufwändigen Untersuchung lächerlich oder befremdlich erscheinen und ihr Ziel verfehlen (Denton, Tsai & Chevrette 1988; Brennan, Hoek & Astridge 1991, zitiert in Klein & Porst 2000: 26).

Nederhof (1983) sowie James & Bolstein (1990) konnten diesen Umstand für sehr kleine prepaid Incentives belegen, die im Rahmen von Untersuchungen mit intensivem Mahnverfahren eingesetzt wurden. In ihren experimentellen Feldstudien nahm die rücklaufsteigernde Wirkung der materiellen Anreize mit jedem zusätzlichen Mahnschreiben ab und verlor sich bis zum Ende der Untersuchung gänzlich. Nach dem Erstanschreiben war gegenüber der Nullvariante noch ein Vorteil zu verzeichnen gewesen¹².

Shettle & Mooney (1999) konnten unter Einsatz eines Incentives von \$ 5 in einer Untersuchung mit einem anspruchsvollen Instrumentarium denselben Effekt einer nachlassenden Wirksamkeit aufzeigen. Nach dem zweiten Mailing besaß das Incentive keine Vorteile gegenüber der Nullvariante mehr. Der nach dem ersten Anschreiben erzielte Vorsprung löste sich in dieser nur aus einem Erinnerungsschreiben bestehenden Untersuchung indessen nur teilweise auf.

Sehr kleine Sachgeschenke und Geldbeträge bzw. moderate Geldbeträge, die im Rahmen einer aufwändigen Untersuchung vergeben werden, dürften sich nach diesen Ergebnissen in Verbindung mit mehreren Mahnaktionen nicht als effektiv erweisen. Sie können bei den Probanden unter Umständen den Eindruck eines ungebührlichen Insistierens angesichts einer Gratifikation erwecken, die dem geforderten Aufwand nicht oder nur in unzureichender Weise entspricht. Eine Berechtigung für ihren Einsatz ergibt sich allenfalls im Rahmen eines reduzierten Verfahrens mit nur einem Mailing. In einem mehrstufigen Mahnverfahren dürfte jedoch nur ein dem Aufwand der Probanden einigermaßen angemessenes Geschenk eine Zunahme des Rücklaufs und einen zusätzlichen Ertrag zur Wirkung der Erinnerungsschreiben gewährleisten. Die Wirkung eines kombinierten Vorgehens mit mehreren Mahnschreiben und materiellen Anreizen scheint damit zumindest zu einem gewissen Grad von der Höhe des Incentives abhängig.

12 In der Studie von Nederhof (1983) wurden Kugelschreiber im Wert von \$ 0.35 mit dem Erstanschreiben verschickt. James & Bolstein (1990) verwendeten prepaid monetäre Anreize im Betrag von \$ 0.25 und \$ 0.50.

Diese Annahme wird durch eine zweite, in diesem Beitrag bereits weiter oben zitierte Studie von James & Bolstein (1992) gestützt, in der sich ein prepaid Incentive von \$ 20 in Kombination mit mehreren Erinnerungsschreiben gegenüber kleinen Beträgen deutlich im Vorteil befand. Zwanzig Dollar erwiesen sich im Hinblick auf die Ausschöpfungsquote als optimal für das gewählte Untersuchungsdesign mit insgesamt vier Kontaktaufnahmen. Im Rahmen eines reduzierten Verfahrens mit nur einem Mailing aber sonst übereinstimmenden Bedingungen hätte aber bereits ein wesentlich billigeres Vorgehen mit \$ 5 einen vergleichbaren Ertrag erbracht. Die \$ 5-Variante wies außerdem unter dem Gesichtspunkt des effizienten Einsatzes von begrenzten Forschungsmitteln deutliche Vorteile auf. Die durch die Verwendung von \$ 20 erzielte Rücklaufsteigerung von acht Prozent gegenüber der kostengünstigeren Variante musste nach vier Anschreiben durch ein Vielfaches an finanziellen Mitteln erkaufte werden.

Es empfiehlt sich beim Einsatz materieller Anreize deshalb nicht nur auf dem Hintergrund von Überlegungen zur optimalen Wirksamkeit einen Testlauf zu organisieren, sondern auch im Hinblick auf die Ermittlung eines in Verbindung mit den übrigen Untersuchungsbedingungen möglichst effizienten und kostengünstigen Verfahrens.

Incentives sollten grundsätzlich nicht losgelöst vom übrigen Kontext einer Untersuchung, sondern wenn immer möglich unter Berücksichtigung aller situativen Merkmale bzw. Kosten- und Nutzenaspekte einer Studie Verwendung finden (Groves et al. 1992; Warriner et al. 1996).

Es kann in der Regel aber davon ausgegangen werden, dass mit Incentives von kleinem bis moderatem materiellem Wert keine Erhöhung der Feldkosten in postalischen und mündlichen Befragungen verbunden ist. Die finanziellen Investitionen zahlen sich meist durch eine zeitliche Verkürzung der Feldphase sowie durch geringere Personal- und Sachkosten aus, da weniger Kontaktaufnahmen mit den Probanden notwendig sind (Groves et al. 2004; Simmons & Wilmot 2004: 7f).

3.3 Incentives und Datenqualität

Es stellt sich die wichtige Anschlussfrage, ob mit dem Gebrauch von Incentives allenfalls Einbußen in der Datenqualität verbunden sind. Wenn dies der Fall wäre, dann würde man sich ihren rücklaufsteigernden Effekt mit qualitativ mangelhaften Daten teuer erkaufen. Beeinträchtigungen in der Datenqualität sind einerseits durch ungünstige Auswirkungen auf die Zusammensetzung der Stichprobe und andererseits durch eine unerwünschte Beeinflussung des Antwortverhaltens der Befragten denkbar.

Auswirkungen auf die Zusammensetzung der Stichprobe

Zur Wechselwirkung von Personmerkmalen und materiellen Anreizen besteht zurzeit noch einiger Bedarf an weiterführender Forschung (vgl. Singer et al. 2000: 187)¹³. Die verfügbaren Ergebnisse erlauben dennoch die einigermaßen zuverlässige Aussage, dass materielle Anreize nicht zu einer Erhöhung des Nonresponsebias führen, indem sie die in sozialwissenschaftlichen Untersuchungen in der Regel bereits übervertretenen oberen Bildungs- und Einkommenssegmente in verstärktem Maße ansprechen (Furse & Stewart 1982; Berry & Kanouse 1987; James & Bolstein 1990; Warriner et al. 1996; Arzheimer & Klein 1998; Shettle & Mooney 1999)¹⁴. Einige Studien vermögen sogar positive Auswirkungen auf die Zusammensetzung der Stichprobe nachzuweisen. So berichten zum Beispiel Nederhof (1983), Singer et al. (1998; 1999; 2000), James (1996, zitiert in Dodd 1998: 24) und Martin et al. (2001) von Untersuchungen, in denen es mit Hilfe von Incentives gelang, die Teilnahmebereitschaft von schlechter gebildeten und einkommensschwachen Personen zu erhöhen. In einer Studie von Martinez-Ebers (1997) ließ sich durch die Vergabe eines prepaid Incentives im Wert von \$ 5 zudem die Beteiligung von in amerikanischen Untersuchungen häufig schlecht vertretenen Probanden lateinamerikanischen Ursprungs steigern. Die Autorin erklärt sich diesen positiven Effekt durch die bei Personen lateinamerikanischer Herkunft im Vergleich zum Rest der US-amerikanischen Bevölkerung größere kulturelle Bedeutung der Norm reziproken Handelns, die mit dem Incentive offenbar aktiviert wurde (vgl. Martinez-Ebers 1997: 83).

Bei Untersuchungsteilnehmern aus der ökonomischen und bildungsmäßigen Unterschicht wird im Gegensatz zu dieser Interpretation oft ein geringerer Grad der Verpflichtung zur Reziprozität und eine größere Empfänglichkeit für die ökonomische Seite von Incentives angenommen (vgl. Gelb 1975: 109; Warriner et al. 1996: 557-558). In jenen Fällen, wo sich materielle Anreize positiv auf die Teilnahmebereitschaft dieser Bevölkerungssegmente

13 Abschließende Antworten werden erst möglich sein, wenn genügend im Detail dokumentierte und unter kontrollierten Bedingungen durchgeführte Studien für Analysen zugänglich sind (vgl. die Bemerkungen von Yu & Cooper 1983; Yammarino et al. 1991; Church 1993: 74; Groves et al. 1999; Singer et al. 2000).

Wenn in der Forschungsliteratur widersprüchliche Ergebnisse berichtet werden, dann ist dies zumindest zum Teil auch auf die ungenügende Datenlage zurückzuführen, die eine Analyse von spezifischen Interaktionseffekten bei gleichzeitiger Kontrolle anderer relevanter Untersuchungsbedingungen nicht oder nur bedingt erlauben.

14 Ein gegenteiliger Befund wird von Reuband (1999) berichtet. In seiner bundesdeutschen Stichprobe führte der Einsatz von Telefonkarten als Incentive zu einer größeren Beteiligung von bildungsnahen und einkommensstarken Personen. Die Telefonkarte wurde allerdings erst im Rahmen der dritten Mahnung verschickt (refusal conversion). Es wurden zudem Telefonkarten verwendet, die durch ihre Bildmotive (Kunst aus verschiedenen Epochen) besonders die besser Gebildeten angesprochen haben dürften (Reuband 1999: 66).

auswirkten, waren auch tatsächlich häufig größere Geldbeträge im Spiel bzw. schien trotz der persönlichen Kosten, die mit der Teilnahme verbunden waren, durch das Incentive ein zusätzlicher materieller Nutzen gegeben (James 1996, zitiert in Dodd 1998: 24; Singer et al. 2000: 185; Martin et al. 2001). Eine vollständige Korrektur des Nonresponsebias ist durch die Vergabe von Incentives freilich nicht zu erwarten. Materielle Anreize können aber je nach kulturellem Hintergrund unterschiedliche teilnahmerelevante Wertsysteme aktivieren (Muir & Weinstein 1962: 538; Shettle & Mooney 1999) und bei differentieller Anwendung möglicherweise ein Mittel zur Reduktion systematischer Ausfälle darstellen.

Auswirkungen auf das Antwortverhalten der Probanden

Nach den bisher vorliegenden Forschungsergebnissen sind durch materielle Anreize nicht gravierende Einbußen in der Antwortqualität zu erwarten.

Incentives haben keine Abstriche in der Sorgfalt und Vollständigkeit, mit der Fragebögen ausgefüllt werden, zur Folge (Furse et al. 1981; Shettle & Mooney 1999; Davern, Rockwood, Sherrod & Stephen 2003), weil sie, wie früher befürchtet, die intrinsische Motivation zu Gunsten eines extrinsischen Verstärkers unterminieren (vgl. Hansen 1980: 78-82). Sie können im Gegenteil zuweilen sogar eine größere Anstrengungsbereitschaft bei der Bearbeitung der Befragungsinstrumente bewirken (James & Bolstein 1990; Willimack, Schuman, Pennell & Lepkowski 1995; Singer, van Hoewyk et al. 1999; Singer et al. 2000). In vielen sozialwissenschaftlichen Studien dürfte das intrinsische Interesse der Befragten zu Beginn der Untersuchung ohnehin eher gering sein (Singer, van Hoewyk & Maher 1998: 154) und den Probanden durch ein angemessenes Incentive der Wert ihrer Teilnahme und die Bedeutung der Studie erst richtig bewusst werden (vgl. Berry & Kanouse 1987: 104-105; Shettle & Mooney 1999: 232)¹⁵.

Es scheinen deshalb vielmehr Bedenken angebracht, dass auf dem Hintergrund einer erhöhten persönlichen Verpflichtung durch Incentives das Antwortverhalten in Richtung sozialer Erwünschtheit beeinflusst werden könnte. Auch diese Bedenken sowie die Befürchtung, Incentives würden sich nachteilig auf die Teilnahmebereitschaft in nachfolgenden Untersuchungen auswirken, weil sie in der Bevölkerung die Erwartung nach regelmäßiger extrinsischer Verstärkung in Form von Geld oder Sachgeschenken entstehen lassen, können nach Durchsicht der empirischen Literatur zumindest vorläufig entschärft werden: Materielle Anreize können zwar dazu führen, dass die Befragten eher in die vom Auftrag-

15 Incentives dürften bei vielen Probanden das Interesse an einer Untersuchung erst richtig wecken und so eine vorzeitige Entscheidung zu deren Ungunsten verhindern. Andere Vorteile (subjektive Nutzenaspekte) der Studie werden dann auf dem Hintergrund der gesteigerten Aufmerksamkeit unter Umständen ebenfalls wahrgenommen und entscheidungsrelevant.

geber gewünschte Richtung antworten (Sponsorship-Effekt: James & Bolstein 1990) oder generell etwas positiver auf die Untersuchungsfragen antworten (Singer et al. 1998; 2000). Zuweilen wird durch sie auch die Erwartung nach Entschädigung in späteren Untersuchungen leicht angehoben (Singer 1998; Singer et al. 1998). Verzerrungen in Richtung sozialer Erwünschtheit (James & Bolstein 1990; Willimack et al. 1995; Martinez-Ebers 1997; Singer et al. 1998; Shettle & Mooney 1999; Singer et al. 2000) oder eine Reduktion der Teilnahmebereitschaft in späteren Untersuchungen bzw. späteren Wellen einer Panelstudie (Arzheimer & Klein 1998; Singer 1998; Singer et al. 1998; Singer, Groves et al. 1999), die keine Vergabe von Incentives vorsahen, wurden bisher jedoch nicht oder nur in vernachlässigbarer Weise festgestellt.

4 Empfehlungen für die Forschungspraxis

Aus den dargestellten Befunden lassen sich einige allgemeine Empfehlungen für die Forschungspraxis ableiten¹⁶. Es gilt dabei zu unterscheiden zwischen Maßnahmen, die sich in der Bundesrepublik und im näheren deutschsprachigen Umfeld in Untersuchungen bewährt haben und solchen, für die zwar genügend Beweiskraft aus dem angelsächsischen Raum vorliegt, die empirische Bewährungsprobe in unseren Breitengraden aber noch weitgehend aussteht.

1. Materielle Anreize können in postalischen, aber auch in mündlichen und telefonischen Befragungen ohne große Bedenken eingesetzt werden. Sie stellen eine bewährte Methode zur Rücklaufsteigerung dar. Eine wichtige Voraussetzung ist allerdings, dass sie gleich beim Erstkontakt, ohne an Bedingungen geknüpft zu sein, gegeben werden.
2. In den meisten Fällen genügen bereits eher symbolisch als finanziell bedeutsame Sachgeschenke oder Geldbeträge (im Sinne einer kleinen Anerkennung, eines kleinen Dankeschöns), um eine deutliche Steigerung der Ausschöpfungsquote zu erreichen. Es ist jedoch darauf zu achten, dass ein Incentive angesichts des von den Befragten verlangten Aufwands nicht als allzu gering und „lächerlich“ wahrgenommen wird.
3. Im deutschen Sprachraum wurden im Rahmen von wissenschaftlichen Befragungen bisher mit Erfolg kleine (oft geldnahe) Sachgeschenke im Sinne von prepaid Incentives eingesetzt. Mit unkonkonditional beim Erstkontakt gegebenen Geldbeträgen wurden bisher noch wenige Erfahrungen gesammelt. Eine große Zahl vor allem amerikanischer Studien zeigt jedoch, dass Geldbeträge im Vergleich zu Sachgeschenken und geldnahen Incentives oft eine größere rücklaufsteigernde Wirkung erzielen.

16 Sehr gute, praxisorientierte Hinweise zum Einsatz von Incentives in postalischen Befragungen finden sich auch bei Stadtmüller & Porst (2005).

4. Mit dem ökonomischen Wert eines monetären Anreizes steigt in der Regel auch der Rücklauf in einer Untersuchung. Diese Tatsache ist für den angelsächsischen Raum wiederum gut belegt, im deutschen Sprachraum hingegen noch wenig untersucht. Wie der Zusammenhang zwischen der Höhe eines Incentives und der Teilnahmebereitschaft genau aussieht und welche Rolle im Rahmen dieser Wechselwirkung andere Elemente des Untersuchungsdesigns spielen, ist jedoch noch nicht restlos geklärt. Es liegt aber nahe, in jeder Untersuchung von einer vom Aufwand für die Befragten abhängigen oberen Grenze der Wirksamkeit monetärer Incentives auszugehen. Die Suche nach der optimalen Höhe eines Incentives erfordert im Prinzip für jede einzelne Studie einen Testlauf. Dabei sind im Hinblick auf eine effiziente Nutzung der vorhandenen Forschungsmittel auch Überlegungen zur Wirtschaftlichkeit anzustellen. Das wirksamste Incentive ist nicht immer gleichzeitig auch das effizienteste.
5. Die Qualität der Daten und die Teilnahmebereitschaft in späteren Untersuchungen scheinen nach dem gegenwärtigen Stand des Wissens durch die Verwendung von materiellen Anreizen nicht gefährdet. Der differentielle Einsatz von Incentives bei unterschiedlichen Bevölkerungsgruppen kann unter Umständen sogar zu einer leichten Reduktion des Nonresponsebias beitragen. Die Belege für diesen letzten Sachverhalt sind jedoch noch spärlich und zum Teil widersprüchlich, so dass weitere Forschungsergebnisse abzuwarten sind.

Literatur

- Adams, J. S. (1965). Inequality in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2) (pp. 267-299). New York: Academic Press.
- Armstrong, S. J. (1975). Monetary incentives in mail surveys. *Public Opinion Quarterly*, 39, 111-116.
- Arzheimer, K. & Klein, M. (1998). Die Wirkung materieller Incentives auf den Rücklauf einer schriftlichen Panelbefragung. *ZA-Information*, 43, 6-31.
- Berry, S. H. & Kanouse, D. E. (1987). Physician response to a mailed survey. *Public Opinion Quarterly*, 51, 102-114.
- Biener, P. M. & Kidd, H. J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychology and Marketing*, 11, 483-492.
- Blasius, J. & Reuband, K.-H. (1996). Postalische Befragungen in der empirischen Sozialforschung. Ausschöpfungsquoten und Antwortqualität. *Planung & Analyse*, 1, 35-41.
- Blau, P. M. (1964). *Exchange and power in social life*. New York: John Wiley and Sons.
- Brehm, J. W. (1966). *A theory of psychological reactance*. New York: Academic Press.

- Brehm, J. W. & Cole, A. (1966). Effect of a favor which reduces freedom. *Journal of Personality and Social Psychology*, 3, 420-426.
- Cavusgil, S. T. & Elvey-Kirk, L. A. (1998). Mail survey response behavior: A conceptualization of motivating factors and an empirical study. *European Journal of Marketing*, 32(11/12), 1165-1192.
- Chebat, J.-C. & Cohen, A. (1993). Response speed in mail surveys: Beware of shortcuts. *Marketing Research*, 5(2), 20-25.
- Church, A. H. (1993). Estimating the effects of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, 57, 62-79.
- Davern, M., Rockwood, T. H., Sherrod, R. & Stephen, C. (2003). Prepaid monetary incentives and data quality in face-to-face interviews. Data from the 1996 survey of income and program participation incentive experiment. *Public Opinion Quarterly*, 67, 139-147.
- Denton, J. J., Tsai, C.-Y. & Chevrette, P. (1988). Effects on survey responses of subjects, incentives, and multiple mailings. *Journal of Experimental Education*, 56(2), 77-82.
- Diaz de rada, V. (2005). The effect of follow-up mailings on the response rate and response quality in mail surveys. *Quality & Quantity*, 39, 1-18.
- Diekmann, A. & Jann, B. (2001). Anreizformen und Ausschöpfungsquoten bei postalischen Befragungen. Eine Prüfung der Reziprozitätshypothese. *ZUMA-Nachrichten*, 48, 18-27.
- Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: John Wiley & Sons.
- Dillman, D. A. (1983). Mail and other self-administered questionnaires. In P. H. Rossi, J. D. Wright & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 359-377). New York: Academic Press.
- Dillman, D. A. (2000). *Mail and internet surveys. The tailored design method*. New York: John Wiley & Sons.
- Dodd, T. (1998). Incentive payments on social surveys: A summary of recent research. *Survey Methodology Bulletin*, 43, 23-27.
- Esser, H. (1986a). Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 38(2), 314-336.
- Esser, H. (1986b). Über die Teilnahme an Befragungen. *ZUMA-Nachrichten*, 18, 38-47.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Fox, R. J., Crask, M. R. & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly*, 52, 467-491.
- Furse, D. H. & Stewart, D. W. (1982). Monetary incentives versus promised contribution to charity: New evidence on mail survey response. *Journal of Marketing Research*, 19, 375-380.
- Furse, D. H., Stewart, D. W. & Rados, D. L. (1981). Effects of foot-in-the-door, cash incentives, and followups on survey response. *Journal of Marketing Research*, 18(4), 473-478.

- Gelb, B. D. (1975). Incentives to increase survey returns: Social class considerations. *Journal of Marketing Research*, 12, 107-109.
- Gouldner, A. W. (1960). The norm of reciprocity. *American Sociological Review*, 25(2), 161-178.
- Goyder, J. (1987). *The silent minority. Nonresponse on sample surveys*. Boulder: Westview Press.
- Groves, R. M., Cialdini, R. B. & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 475-495.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2004). *Survey methodology*. John Wiley & Sons, Inc.
- Groves, R. M., Singer, E. & Corning, A. D. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *Public Opinion Quarterly*, 64(3), 299-308.
- Groves, R. M., Singer, E., Corning, A. D. & Bowers, A. (1999). A laboratory approach to measuring the effects on survey participation of interview length, incentives, differential incentives, and refusal conversion. *Journal of Official Statistics*, 15(2), 251-268.
- Hansen, R. A. (1980). A self-perception interpretation of the effect of monetary and non-monetary incentives on mail survey respondent behavior. *Journal of Marketing Research*, 17(1), 77-83.
- Harkness, J., Mohler, P., Schneid, M. & Christoph, B. (1998). Incentives in two German mail surveys 1996/97 & 1997. In A. Koch & R. Porst (Eds.), *Nonresponse in survey research. ZUMA-Nachrichten Spezial 4*. (pp. 201-218). Mannheim: ZUMA.
- Hippler, H.-J. & Seidel, K. (1985). Schriftliche Befragung bei allgemeinen Bevölkerungstichproben – Untersuchungen zur Dillmanschen “Total Design Method”. *ZUMA-Arbeitsbericht*, 16, 39-56.
- Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt, Brace and World.
- Hubbard, R. & Little, E. L. (1988). Promised contributions to charity and mail survey responses. Replication with extension. *Public Opinion Quarterly*, 52, 223-230.
- James, J. M. & Bolstein, R. (1990). The effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys. *Public Opinion Quarterly*, 54, 346-361.
- James, J. M. & Bolstein, R. (1992). Large monetary incentives and their effect on mail survey response rates. *Public Opinion Quarterly*, 56, 442-453.
- Kaldenberg, D. O., Koenig, H. F. & Becker, B. W. (1994). Mail survey response rate patterns in a population of the elderly. Does response deteriorate with age? *Public Opinion Quarterly*, 58, 68-76.
- Klein, S. & Porst, R. (2000). Mail Surveys. Ein Literaturbericht. *ZUMA-Technischer Bericht*, 10/2000.

- Lynn, P. (2001). The impact of incentives on response rates to personal interview surveys: Role and perceptions of interviewers. *International Journal of Public Opinion Research*, 13(3), 326-336.
- Martin, E., Abreu, D. & Winters, F. (2001). Money and motive: Effects of incentives on panel attrition in the survey of income and program participation. *Journal of Official Statistics*, 17(2), 267-284.
- Martinez-Ebers, V. (1997). Using monetary incentives with hard-to-reach populations in panel surveys. *International Journal of Public Opinion Research*, 9(1), 77-86.
- McConaghy, M. & Beerten, R. (2003). Influencing response on the family resources survey by using incentives. *Social Survey Methodology Bulletin*, 51, 27-35.
- Muir, D. E. & Weinstein, E. A. (1962). The social debt: An investigation of lower-class and middle-class norms of social obligation. *American Sociological Review*, 27, 532-539.
- Nederhof, A. J. (1983). The effects of material incentives in mail surveys: Two studies. *Public Opinion Quarterly*, 47, 103-111.
- Porst, R. (1996). Ausschöpfungen bei sozialwissenschaftlichen Umfragen. Die Sicht der Institute. *ZUMA-Arbeitsbericht*, 96/07.
- Porst, R. (1999). Thematik oder Incentives? Zur Erhöhung der Rücklaufquoten bei postalischen Befragungen. *ZUMA-Nachrichten*, 45, 72-87.
- Porst, R., Ranft, S. & Ruoff, B. (1998). Strategien und Maßnahmen zur Erhöhung der Ausschöpfungsquoten bei sozialwissenschaftlichen Umfragen. Ein Literaturbericht. *ZUMA-Arbeitsbericht*, 98/07.
- Reuband, K.-H. (1999). Telefonkarten als Incentives für nicht-kooperative Zielpersonen in postalischen Befragungen. Auswirkungen auf die Teilnahmebereitschaft und die Zusammensetzung der Befragten. *Planung & Analyse*, 3, 63-66.
- Reuband, K.-H. & Blasius, J. (1996). Face-to-Face-, telefonische und postalische Befragungen: Ausschöpfungsquoten und Antwortmuster in einer Großstadt-Studie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 48, 296-318.
- Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen*. Opladen: Leske + Budrich.
- Shettle, C. & Mooney, G. (1999). Monetary incentives in U.S. government surveys. *Journal of Official Statistics*, 15(2), 231-250.
- Simmons, E. & Wilmot, A. (2004). Incentive payments on social surveys: A literature review. *Social Survey Methodology Bulletin*, 53, 1-11.
- Singer, E. (1998). Incentives for survey participation: Research on intended and unintended consequences. *ZUMA-Arbeitsbericht*, 42, 7-29.
- Singer, E., Groves, R. M. & Corning, A. D. (1999). Differential incentives. Beliefs about practices, perceptions of equity, and effects on survey participation. *Public Opinion Quarterly*, 63, 251-260.

- Singer, E., van Hoewyk, J., Gebler, N., Raghunathan, T. & McGonagle, K. (1999). The effect of incentives on response rates in interview-mediated surveys. *Journal of Official Statistics*, 15(2), 217-230.
- Singer, E., van Hoewyk, J. & Maher, M. P. (1998). Does the payment of incentives create expectation effects? *Public Opinion Quarterly*, 62, 152-164.
- Singer, E., van Hoewyk, J. & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, 64, 171-188.
- Stadtmüller, S. & Porst, R. (2005). Zum Einsatz von Incentives bei postalischen Befragungen. *ZUMA How-to-Reihe*, 14.
- Thibaut, J. W. & Kelley, H. H. (1959). *The social psychology of groups*. New York: John Wiley and Sons.
- Thoma, M. & Zimmermann, M. (1996). Zum Einfluß der Befragungstechnik auf den Rücklauf bei schriftlichen Umfragen - Experimentelle Befunde zur „Total-Design-Methode“. *ZUMA-Nachrichten*, 39, 141-159.
- Trussell, N. & Lavrakas, P. J. (2004). The influence of incremental increases in token cash incentives on mail survey response. Is there an optimal amount? *Public Opinion Quarterly*, 68(3), 349-367.
- Warriner, K., Goyder, J., Gjertsen, H., Hohner, P. & Mc Spurren, K. (1996). Charities, no; lotteries, no; cash, yes. Main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly*, 60, 542-562.
- Willimack, D. K., Schuman, H., Pennell, B.-E. & Lepkowski, J. M. (1995). Effects of a prepaid nonmonetary incentive on response rates and response quality in a face-to-face survey. *Public Opinion Quarterly*, 59, 78-92.
- Yammarino, F. J., Skinner, S. J. & Childers, T. L. (1991). Understanding mail survey response behavior. A meta-analysis. *Public Opinion Quarterly*, 55, 613-639.
- Yu, J. & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, 20(1), 36-44.

Korrespondenzadresse

Fred Berger, lic. phil.

Pädagogisches Institut der Universität Zürich

Freie-Str. 36

CH-8032 Zürich

E-mail: fberger@paed.unizh.ch

COMPUTERUNTERSTÜTZTE VERCODUNG DER INTERNATIONAL STANDARD CLASSIFICATION OF OCCUPATIONS (ISCO-88): VORSTELLEN EINES INSTRUMENTS

JÜRGEN H.P. HOFFMEYER-ZLOTNIK, DORIS HESS & ALFONS J. GEIS

Die „International Standard Classification of Occupations“ von 1988 (ISCO-88) stellt das Instrument dar, das für die Harmonisierung der Klassifikation von Berufen in der international vergleichenden Forschung verwendet wird. ISCO-88 ist ein hoch komplexes Instrument, das etwa 30.000 Berufe in 390 Kategorien für berufliche Tätigkeiten abbildet. Vorgestellt wird das Ergebnis der Entwicklung eines für Umfragen in Deutschland einsetzbaren Instrumentariums zur computerunterstützten Feldverschlüsselung der Berufe nach ISCO-88.

The 1988 „International Standard Classification of Occupations“ ISCO-88) is an instrument used to harmonize the measurement of occupations in international research. ISCO-88 is a complex instrument which maps 30,000 different job titles into 390 occupational categories. This article reports the results of our research: an instrument for generating ISCO categories during computer-assisted data collection in Germany.

1 Einführung

Die „International Standard Classification of Occupations“ von 1988 (ISCO-88) ermöglicht eine Klassifikation der national auftretenden Berufe für den internationalen Vergleich. Dieses Instrument kondensiert die etwa 30.000 in Deutschland vorkommenden Berufe auf 390 Codes auf der vierten hierarchischen Ebene. In dem hoch komplexen System von ISCO-88 werden hierbei Definitionsmerkmale für Tätigkeiten berücksichtigt. Die für den nationalen deutschen Gebrauch übliche „Klassifizierung der Berufe“ (KlDB) des deutschen Statistischen Bundesamtes von 1992 strukturiert etwa 29.500 Berufsbenennungen in einer vierstufigen Hierarchie. ISCO-88 und KlDB-92 bauen auf unterschiedlichen Logiken auf. Sie unterscheiden sich vor allem darin, dass bei der ISCO-88 auf der vierten Ebene 390 Definitionen von Berufen stehen und bei der KlDB-92 auf der vierten, der untersten Ebene, eine Auflistung von 29.500 Berufsbezeichnungen erfolgt. Definitionen von Berufen stellen eine härtere Anforderung an die Codierung.

Computerunterstützt kann man die 29.500 Berufsbezeichnungen der KldB-92 im Hintergrund speichern und diese mit der Dateneingabe im Klartext, sofern dieser einer der 29.500 gespeicherten Bezeichnungen entspricht, abgleichen und automatisch vercoden. Allerdings unterliegen Berufsbezeichnungen einem sehr schnellen Wandel. Im Gegensatz dazu weisen Definitionen von Berufen nur einen geringen Interpretationsspielraum auf und sind relativ zeitlos. Sie lassen sich allerdings nicht über Berufsbezeichnungen zuordnen, da in eine entsprechende Definition Tätigkeitsbeschreibungen eingehen, die nicht unbedingt mit Berufsbezeichnungen verbunden sind: Ausbildungsvoraussetzung, Personalführung, Autonomie in der Tätigkeit, „blue-collar“ (gewerblich) mit/ohne Maschinenbedienung, oder „white-collar“ (Dienstleistung, Verwaltung) mit/ohne Kundenkontakt.

Tabelle 1 Struktur der ISCO-88

Berufshauptgruppe		Anzahl Einheiten		
		Berufsgruppe	Berufsuntergruppe	Berufsgattung
1	gesetzgebende Körperschaften, leitende Verwaltungsbedienstete, Geschäftsleiter	3	8	33
2	Wissenschaftler	4	18	55
3	Techniker und gleichrangige nicht-technische Berufe	4	21	73
4	Bürokräfte, kaufmännische Angestellte	2	7	23
5	Dienstleistungsberufe, Verkäufer	2	9	23
6	Fachkräfte in Landwirtschaft und Fischerei	2	6	17
7	Handwerks- und verwandte Berufe	4	16	70
8	Anlagen- und Maschinenbediener	3	20	70
9	Hilfsarbeitskräfte	3	10	25
0	Streitkräfte	1	1	1
	gesamt	28	116	390

ILO 1990: 3

2 Fragestellung

Eine alle benötigten Informationen umfassende Abfrage von Berufen bzw. beruflichen Tätigkeiten ist das eine. Die Vercodung dieser Informationen nach der ISCO-88 ist das andere. Eine maschinelle Vercodung ist nur in Teilen möglich. Dieses bedeutet bisher eine Vercodung großer Teile per Hand. Eine Vercodung per Hand ist jedoch kostenaufwendig. Daher wird der Weg einer weitgehend maschinellen Vercodung gesucht, ein Weg, der bei computerunterstützten Befragungstechniken möglich sein sollte.

Ein erster Weg ist die Anwendung des vom Statistischen Bundesamtes entwickelten Umsteigeschlüssels für den Umstieg von den Berufsbezeichnungen der KldB-92 zu den Berufsdefinitionen der ISCO-88. Allerdings sind Berufsbezeichnungen nicht unbedingt mit Tätigkeitsbeschreibungen verbunden. Bei der KldB-92 geschieht die Differenzierung vom Speziellen zum Generellen, indem die Berufsbezeichnungen erfasst und diese dann zu Gruppen von Berufsbezeichnungen, nicht von Tätigkeitsbeschreibungen, zusammengefasst werden. Ein von A.J. Geis und J. H.P. Hoffmeyer-Zlotnik (2001: 125 ff.) beschriebener empirischer Test einer automatischen Umsetzung von KldB-92 auf ISCO-88 zeigt, dass von ca. 12.000 Berufsangaben nur 54 % der nach KldB-92 vercodeten Berufe eindeutig einem ISCO-88-Code auf der vierten Hierarchieebene genau zuzuordnen waren. Benutzt man die Regeln des Statistischen Bundesamtes für die Zuordnung der KldB-92-Codes zu den ISCO-88-Codes, so ergibt sich auf der vierten Stelle eine Zuordnungsgenauigkeit von 65,3 % (Geis & Hoffmeyer-Zlotnik 2001: 131, 134). Das restliche Drittel der Berufsangaben eines Surveys wird unpräzise oder falsch zugeordnet. Aus diesem Grund ist für eine computerunterstützte Vercodung der ISCO-88 ein Ansatz zu wählen, der unabhängig von der KldB-92 und deren Logik, von den Berufsbezeichnungen auszugehen, vorgehen muss.

3 Umsetzung

Das Projektteam, bestehend aus Jürgen H.P. Hoffmeyer-Zlotnik und Alfons J. Geis vom ZUMA sowie Doris Hess vom infas, Institut für angewandte Sozialwissenschaft, testete zunächst in einem Vortest, welche Zusatzinformationen so zentral sind, dass sie unbedingt Berücksichtigung finden müssten (siehe Hoffmeyer-Zlotnik, Hess & Geis 2004: 32-47). Das Ergebnis: Auf der Ebene der Berufshauptgruppen ist für relevante Einzelvariablen die Zuordnung von Einzelmerkmalen zu Beruf teils sehr gut. Etwa zwei Drittel der Zuordnungen waren korrekt. Das gilt allerdings primär für drei Berufshauptgruppen: die Leitung (1), die Akademiker (2) und die Handwerker (7). Auf der Ebene der zweistelligen Berufsgruppen waren teilweise auch noch gute Zuordnungsquoten über Variablenkombinationen möglich. Auf der dreistelligen Ebene der Berufsuntergruppen war im Vortest die Höhe der Trefferquote einer richtigen Zuordnung nur noch für wenige Fälle akzeptabel. Auf der Ebene der vierstelligen Berufsgattungen brachte das Set an Zusatzinformationen wenig Präzisierung bei der automatischen Vercodung von Berufsangaben in den ISCO-Kategorien. Das Fazit war: Es lässt sich über eine mehrstufige Berufsabfrage, die z.B. mit der ‚Stellung im Beruf‘ oder der ‚Bildung‘ beginnt, das Drittel der über die Berufsbezeichnung maschinell codierbaren Nennungen nicht wesentlich erhöhen. Dieses zeigte uns, „dass die Nennung der Berufsangaben über ein vorgegebenes Kategoriensystem gesteuert werden muss. Die offene Antwort der Berufsnennungen erfordert zur Vercodung

die komplexe Struktur des menschlichen Gehirns und nicht die einfache Struktur der einnull-Logik des derzeitigen Computers." (Hoffmeyer-Zlotnik, Hess & Geis 2004: 47).

In den folgenden Tests wendeten wir uns von der Idee ab, Logik der manuellen Vercodung in den Computer zu übertragen. Die Logik der manuellen Vercodung ordnet eine Berufsbeneennung über nacheinander heranzuziehende Zusatzinformationen ein, analog denen, die bei Hoffmeyer-Zlotnik, Hess und Geis (2004) aufgelistet wurden (wie „Klassifikation der beruflichen Stellung“, „Bildungsniveau“, „berufliche Ausbildung“, „Branche“, „Arbeitsgerät“, „Art der Arbeit mit Maschinen“ und für den Beruf „zentrale Tätigkeiten“). Die neue Idee geht von einem geschlossenen Kategoriensystem aus, das dann von einer Oberkategorie in eine hierarchische Baumstruktur einmündet. Dieses Kategoriensystem beginnt mit einer Einstiegstabelle, über die die Grundverortung der Berufe zu geschehen hat. Über bis zu drei weitere Hierarchieebenen wird dann der Befragte in die Feinverortung zum zutreffenden ISCO-Code auf der Ebene der vierstelligen Berufsgattungen geführt. Der Erfolg ist sichtbar, obwohl wir den schwierigsten Modus gewählt hatten, eine mündlich-telefonische Befragung, bei der es nicht möglich ist, die Befragten mit visuellen Hilfen zu unterstützen. Dennoch haben wir in einer ausführlichen Testserie, die uns „infas, Institut für angewandte Sozialwissenschaft“, Bonn, ermöglicht hat, den Anteil der zutreffend eingeordneten Fälle kontinuierlich erhöhen können.

Hierzu wurden die Berufsangaben einerseits in hierarchischer Abfolge nachvollziehbar im Feld vercodet, andererseits aber auch für alle Befragten offen erfasst, so dass nach Feldende eine manuelle Vercodung durch den Stab der auf die Berufsvercodung nach ISCO-88 trainierten Verncoder von ZUMA durchgeführt werden konnte. Damit war sowohl ein Nachvollziehen der Antwortfindung durch die Befragten möglich als auch ein Abgleich mit den Zuordnungen der Berufe durch die ZUMA-Verncoder, denen die Selbstzuordnung der Befragten nicht vorlag.

Im letzten Test erreichten wir einen Anteil an zutreffend von den Befragten vercodeten Berufsangaben, der unsere eigenen Erwartungen übertroffen hat (siehe Tabelle 2).

Tabelle 2 Zutreffende Zuordnung durch den Befragten

Zeitpunkt	Fälle	zutreffend	zutreffend, ohne: weiß nicht
Mai 2004	197	36 %	
März 2005	194	55 %	58 %
August 2005	203		72 %
November 2005	119	76 %	84 %

Quellen: infas ISCO-Piloten 05/04, 03/05, 08/05, 11/05

Das Instrument wurde nach der Testphase in einer großen Umfrage getestet und damit dem „Härtetest“ unterzogen:

Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg hat infas Anfang September 2005 mit der Durchführung einer repräsentativen Querschnittsbefragung im Rahmen des Projektes „Zusammenlegung von Arbeitslosen- und Sozialhilfe mit Einführung des SGB II: Übergangsmuster, Strukturen und Folgen für die Betroffenen“ beauftragt. Im Rahmen dieser Querschnittserhebung wurden in der Zeit von November 2005 bis März 2006 21.000 Interviews mit einer durchschnittlichen Interviewdauer von 80 Minuten bei erwerbsfähigen Hilfebedürftigen telefonisch oder persönlich mündlich durchgeführt. Die Befragung selbst ist als kombinierte Querschnitts- und biographische Längsschnittstudie angelegt und wurde computerunterstützt (CATI/CAPI) durchgeführt. Die Berufsbiographie wurde retrospektiv erfasst, wobei die beruflichen Tätigkeiten erstmals computerunterstützt nach diesem hier vorgestellten neuen Verfahren erhoben wurden. Insgesamt wurden in der SGB II-Querschnittserhebung 42.055 Angaben zur beruflichen Tätigkeit erfasst, davon liegen für 5.778 weiß-nicht-Angaben die offenen Textangaben vor.

Tabelle 3 Erhaltene ISCO-Angaben auf den unterschiedlichen Ebenen

Ebene der ISCO-Vercodung	Angaben
Tätigkeiten zur Vercodung insgesamt	42.055
davon:	
– Ebene „Berufsgattung“ (4-stellig)	34.889
– Ebene „Berufsuntergruppe“ (3-stellig)	683
– Ebene „Berufsgruppe“ (2-stellig)	250
– Ebene „Berufshauptgruppe“ (1-stellig)	241
– weiß-nicht-Antworten mit offenen Textangaben	5.778
– verweigert	214

Quelle: IAB: SGB II-Querschnittserhebung 2006

Nach Sichtung dieser offenen Texte ist es aus unserer Sicht möglich, gut zwei Drittel der offenen Textangaben ohne manuelle Vorarbeit maschinell über ein Wörterbuch zu vercoden.

4 Instrumentenbeschreibung

Das Herzstück einer computerunterstützten Feldvercodung von ISCO-88 ist die Einstiegstabelle. Unsere Einstiegstabelle umfasst neben den Kategorien für „weiß nicht“ und „verweigert“ 39 inhaltliche Kategorien¹, von denen einige einstellig, die Mehrheit zweistellig aber auch einige Exoten, die schwer über eine Baumstruktur zu erreichen sind, vierstellig sind.

Einstellig über den Berufshauptgruppencode steigen die Fachkräfte in Landwirtschaft und Fischerei (6), die Handwerker (7) und die Anlagen- und Maschinenbediener (8) ein. Landwirte und Fischer werden, unterschieden nach ihrer Produktion für den Markt- oder den Eigenbedarf, über zwei weitere Ebenen der Abfrage zur vierstelligen Berufsgattung geführt. Die Handwerker werden zunächst in einer breiten Auflistung mit wenigen Ausnahmen über den Code 7 gesammelt und dann vom Einsteller über den Zweisteller auf den Dreisteller und über diesen auf den Viersteller geführt. Die Ausnahmen stellen jene Personen dar, die sich nicht unbedingt über die Oberkategorie einfangen lassen, wie z.B. der „gelernte Gebäudereiniger“ (Code 7143) oder die „Metallarbeiter, Mechaniker, Monteure, Werkzeugmacher“ (Code 72). Die Anlagen- und Maschinenbediener (Code 8) werden ebenfalls in einer breiten Aufzählung des heterogenen Feldes gesammelt und auf den Zweisteller, die Exoten, bestehend aus dem „Deckspersonal auf Schiffen“ (Code 8340), sofort auf den Viersteller geführt. In der Einstiegstabelle wird bei dieser Gruppe nur für die, die in der „Gewinnung von Rohprodukten“ (Code 81) beschäftigt sind, noch einmal nachgefasst. Da sich unter Anlagen- und Maschinenbediener auch Personen aus den Handwerker-Kategorien (Codes 71 und 72) einordnen, muss mit Querverweisen zu den Listen der „Handwerker“ (Code 7) gearbeitet werden. Von der zweistelligen geht es über die dreistellige auf die vierstelligen Ebene der Anlagen- und Maschinenbediener.

Für den Befragten weniger nachvollziehbar strukturiert sind die Berufshauptgruppen der „Techniker und gleichrangigen nichttechnischen Berufe“ (Code 3). Diese sind zunächst von den Akademikern (Code 2) und von den „Bürokräften, kaufmännischen Angestellten“ (Code 4) zu unterscheiden. Daher werden auf der Einstiegstabelle zunächst die Bürokräfte (Code 4) abgesondert, dann die Techniker in den zweistelligen Berufsgruppen gesammelt und, abgesehen von den „Nicht-wissenschaftlichen Lehrkräften“ (Code 33), die Akademiker erfasst. Für einzelne, nicht über die Berufsgruppen erreichbare Berufsfelder, muss schon in der Einstiegstabelle über die dreistellige Berufsuntergruppe, wie z.B. für die „Bediener optischer oder elektronischer Anlagen“ (Code 313) oder „Schiffs- oder Flugzeugführer“ (Code 314) oder gar über die vierstelligen Berufsgattung bei „Baukontrolleur“ (Code 3151) und „Sicherheitskontrolleur“ (Code 3152), der Befragte richtig geleitet wer-

1 Die Zahlen in Klammern stellen die zutreffenden ISCO-88-Codes dar.

den. Die Weiterleitung von der zweistelligen Berufsgruppe zur vierstelligen Berufsgattung ist dann unproblematisch.

Bei den Bürokräften (Code 4) sind lediglich die mit Kundenkontakt (Code 42) und die ohne Kundenkontakt (Code 41) zu unterscheiden. Danach geht es über den Zweisteller und den Dreisteller hierarchisch zum Viersteller. Bei den „Dienstleistungsberufen und Verkäufern“ (Code 5) beginnt das Sammeln der Betroffenen in der Einstiegstabelle ebenfalls über die zweistelligen Berufsgruppen mit einer tieferen Gliederung bei denen, die sich nicht ohne weiteres als „Dienstleistungsberuf“ verstehen, z.B. der „Arzt-, Zahnarzt-helferin“ (Code 513).

Ebenfalls einfach in einer zweistelligen Erfassung über die Einstiegstabelle und einem hierarchischen Weiterleiten sind die „Hilfskräfte“ (Code 9) zu verorten. Bei dieser Gruppe müssen jedoch auf der Ebene der Berufsgattung Kategorien über die Definition der Gattung erheblich erweitert werden, z.B. „Hilfskräfte oder Reinigungspersonal in Büros, Hotels oder sonstigen Einrichtungen, auch Küchenhilfe in Küchen, Hotels oder sonstigen Einrichtungen“ (Code 9132).

„Angehörige gesetzgebender Körperschaften, leitende Verwaltungsbedienstete und Führungskräfte in der Privatwirtschaft“ (Code 1) lassen sich über die Berufsgruppen in der richtigen Untergliederung am Einstieg erfassen und dann hierarchisch weiterleiten. Die Akademiker (Code 2) erfordern nicht nur einen Einstieg über die Berufsgruppen, sondern bei den „sonstigen Wissenschaftlern“ (Code 24) einen Einstieg über die Berufsuntergruppen.

5 Probleme

Ein zentraler Ordnungsfaktor der Einstiegstabelle ist die Reihenfolge der Anordnung der Kategorien. Dieses ist besonders wichtig bei einer Liste, die am Telefon keine visuelle Unterstützung ermöglicht. Da die Liste sehr lang ist, muss der Befragte die Möglichkeit haben, dann, wenn die für ihn zutreffende Kategorie erreicht ist, die Einstiegstabelle zu verlassen. Dieses bedeutet, dass Fehlzuordnungen nur vermieden werden können, wenn Personen, die im Bereich der Berufsgruppen in ihrer Selbstwahrnehmung mehr als eine Möglichkeit der Zuordnung haben, bereits über die richtige Kategorie abgeleitet wurden, bevor ein Zweifel aufkommen kann. Damit müssen zunächst die Bürokräfte (4) identifiziert und abgeleitet werden, dann der Handel auf den unterschiedlichen Stausebenen verortet werden, wie z.B. „Fachfrau, -mann in Handel“ (34), „Verkäufer in Geschäften oder auf Märkten“ (52) und „Straßenhändler“ (911). Dort, wo eine Unterscheidung zwischen „gelernt“ (Berufshauptgruppen 5 und 7) und „angelernt“ (9) notwendig ist, ist der niedrige Status vor dem hohen zu erfassen.

Ein weiterer Ordnungsfaktor der Einstiegstabelle, der diese stark aufbläht, sind jene Positionen, die unter einem Etikett von Berufshauptgruppe oder Berufsgruppe nicht sofort sichtbar werden, wie die „Schiffs- und Flugzeugführer“ (314) als Untereinheit von „Fachkräften in Physik und Ingenieurwissenschaften“ (31) oder wie der „gelernte Gebäudereiniger“ (7143) als Untereinheit zu „Bauberufe“ (71). Entsprechende Positionen müssen, obwohl unterhalb der Ebene der zweistelligen Berufsgruppe, separat abgefragt werden.

Ein dritter Punkt, der bei der Erstellung der Einstiegstabelle zu beachten ist, ist einerseits die teils sehr detaillierte Aufzählung von Merkmalen oder Berufen, die unter einer Kategorie abzuhandeln sind, z.B. die „Handwerker oder in einem verwandten gelernten Beruf, auch Bergbau, Bauberufe, Innenausbau, Kunsthandwerk, Druck, Textil, Bekleidung und Nahrungsmittel, auch Bäcker, Metzger, Schneider, Schuster usw.“ (7). Andererseits erfordern manche Positionen auch eine definierende Beschreibung wie z.B. die „nicht-wissenschaftliche Lehrkraft ohne Hochschulabschluss für schulische und berufliche Aus- und Weiterbildung, auch in Kindergarten oder Hort“ (33).

Ein weiterer Problempunkt, auf den oben schon eingegangen wurde, sind jene Personen, die sich zunächst auf einer oberen Ebene falsch zuordnen und auf einer Ebene tiefer dann wieder in ihren Bereich hinüber geleitet werden müssen, wie z.B. Personen, die sich zunächst unter „Anlagen- und Maschinenbediener“ (8) einordnen, jedoch zur Kategorie der „Handwerker“ gehören.

6 Fehleranalyse

Bei der Fehleranalyse, d.h. dem Vergleich der Selbstzuordnung der Befragten mit der Fremdzuzuordnung durch die Vercoder, treten zwei Arten von Fehlern auf: Entweder der Vercoder sieht den Befragten anders, als dieser sich selbst einordnet, oder der Befragte steigt über einen falschen Ausgangscode ein und läuft in die Irre.

Aus diesen zwei Arten von Fehlern entstehen drei Situationen (siehe Tabelle 4):

Tabelle 4 Vergleich Selbstzuordnung Befragte und Fremdzuzuordnung Vercoder

Art der Abweichung		Prozent
1	Befragter falsch zugeordnet, dann Abbruch mit „weiß nicht“	10
2	Befragter falsch zugeordnet, Vercoder plausibel	14
3.1	Befragter plausibel verortet, Vercoder nach unten*) abweichend	9
3.2	Befragter plausibel verortet, Vercoder nach oben*) abweichend	10
--	Befragter plausibel verortet, Vercoder übereinstimmend	56

*) „unten“ bedeutet niedrigere Bewertung und höhere Codezahl, „oben“ bedeutet höhere Bewertung und niedrigere Codezahl

1. Der Befragte steigt über einen falschen Ausgangscode ein, findet sich nicht zurecht und bricht ab. In diesem Fall muss eine richtige Zuordnung über eine Nachvercodung gefunden werden.
2. Der Befragte ordnet sich falsch zu, ein Vercoder könnte dieses korrigieren.
3. Der Vercoder ordnet den Befragten falsch zu, obwohl der Befragte sich selbst richtig verortet. Der Vercoder würde aufgrund fehlender Informationen den Fehler machen.

Für den ersten Fehlertyp sollen exemplarisch zwei Beispiele angeführt werden:

Der Befragte gibt als Beruf „Werkzeugprüfer“ an.

Der Befragte sieht sich als „Handwerker“ und steigt über den

Code 72: „Metallarbeiter ..., Werkzeugmacher“ ein. Es folgt der

Code 721: „Former für Metallguss“. An dieser Stelle bemerkt der Befragte, dass er sich in einer Sackgasse befindet und bricht mit „weiß nicht“ ab. Jetzt kann der Vercoder tätig werden:

Der „Werkzeugprüfer“ ist per Definition ein „Sicherheitskontrolleur“, Code 3152. Wir holen ihn in der Einstiegstabelle direkt ab, da er sich über Code 31: „Ingenieurwissenschaftlicher Techniker“ einsteigend nie zuordnen würde.

Der Befragte gibt „Melker“ an.

Der Befragte reagiert auf das beim Vorlesen der Einstiegstabelle erste Auftreten des Begriffes „Agrar...“ und steigt über den

Code 32: „Medizinische Gesundheitskraft, ..., Agrartechniker“ ein. Es folgt

Code 321: „Biotechniker“. Der Befragte merkt, dass er sich in einer Sackgasse befindet und bricht mit „weiß nicht“ ab. Auch hier kann der Vercoder tätig werden:

Weiter unten auf der Einstiegstabelle folgt der Code 6: „Fachkraft ..., in der Landwirtschaft“. Über Code 612: „Tierwirtschaftlicher Beruf (für die Marktproduktion)“ gelangt man zu Code 6121: „Milchviehalter“, worunter auch, nicht-selbständig, der „Melker“ zu verorten ist.

Für den zweiten Fehlertyp, die falsche Zuordnung durch den Befragten, die nur durch eine Nachvercodung zu bereinigen ist, sollen zwei Beispiele angeführt werden:

Der Befragte gibt als Beruf „Bäcker“ an.

Der Befragte sieht sich als „Dienstleister“ und steigt über den

Code 51: „im Dienstleistungsberuf“ ein. Es folgt der

Code 512: „hauswirtschaftlicher Bereich oder Gastgewerbe“. Danach bleibt nur noch Code 5122: „Koch“, da alle anderen Tätigkeiten noch weniger zutreffen.

Es wäre besser gewesen, der Befragte hätte eingesehen, dass er keinen zutreffenden Code findet, und reagiert mit „weiß nicht“ spätestens auf der Stufe des letzten Codes. Der Code 7412: „Bäcker“ startet mit dem Code 7 für „Handwerker“.

Der Befragte gibt als Beruf, zweistufig beantwortet, für Tätigkeit an: „Gewinde bohren in Teilen“, und auf Nachfrage nach dem Namen des Berufes: „Teilefertiger für Großserien, Schlosser“.

Der Befragte sieht sich als „Handwerker“ und steigt über den

Code 72 „Metallarbeiter“ ein. Es folgt der

Code 723 „Maschinenmechaniker oder -schlosser“. Es folgt

Code 7233 „Landmaschinen- oder Industrieschlosser“.

Der Beruf ist falsch zugeordnet, da der Befragte kein „Handwerker“ ist und damit nicht über Code 7 einsteigen darf, sondern „Maschinen- und Anlagenbediener“ ist, und damit mit Code 8 einsteigen müsste, um dann zum Code 8211: „Werkzeugmaschinenbediener“ zu gelangen. Der Befragte stieg in die Selbstzuordnung über die Berufsbezeichnung ein, er hätte über die Tätigkeit einsteigen müssen. Die Tätigkeit besteht daraus, eine Maschine zu bedienen.

Die Befragte gibt als Beruf, zweistufig beantwortet, für Tätigkeit an: „Elektroteile an der Maschine“, und auf Nachfrage nach dem Namen des Berufes: „Montiererin“.

Die Befragte sieht sich als „Metallarbeiterin“ im „Handwerk“ und steigt über den

Code 72: „Metallarbeiter, ... Monteure“ ein. Es folgt

Code 724: „Elektromonteur“, gefolgt auf der dritten Ebene von

Code 7242: „Elektronikmonteur“.

Die Tätigkeit mag richtig sein, die Zuordnung jedoch nicht, denn die Befragte ist von der Ausbildung her „ungelernt“ und von der Stellung im Beruf her in „einfacher Tätigkeit“. Die Befragte hätte über Code 93: „Hilfsarbeiter im verarbeitenden Gewerbe“ zu Code 9321: „Montagehilfsarbeiter“ kommen müssen. Diese Fehlzugeordnung ist nur von einem erfahrenen Vercoder, der auch die Kontrollvariablen berücksichtigt, zu entdecken.

Der Fehlertyp drei: Der Befragte selbst ordnet sich richtig ein, er wird allerdings vom Vercoder falsch zugeordnet. Dieser Fehlertyp, der bei jeder ISCO-Vercodung geschehen kann, könnte über eine Kontrollvercodung aufgedeckt werden. Da die Intercoderreliabilität bei einem guten Vercoderstab nur Abweichungen innerhalb einer kleinen Toleranz aufweist, bleiben solche Fehler unentdeckt.

Die Befragte gibt als Beruf, zweistufig beantwortet, für Tätigkeit an: „Werkzeugkonstruktion“, und auf Nachfrage nach dem Namen des Berufes: „Konstrukteur“.

Die Befragte sieht sich als „Technikerin“ und steigt über den

Code 31: „Ingenieurwissenschaftlicher Techniker“ ein. Es folgt

Code 3118: „Technischer Zeichner“.

Der Vercoder steigt über das „Handwerk“ (Code 7) ein und verortet die Befragte unter Code 7222: „Werkzeugmacher“. Begründet wurde dieses über einen mittleren Schulabschluss und eine nicht all zu hohe Autonomie im Job.

Weniger problematisch sind Zuordnungen, die auf der Ebene von Status oder Prestige nichts verändern. Das nachfolgende Beispiel zeigt allerdings, dass die zunächst missverständlich benannte Tätigkeit dem Vercoder als Grundlage diente, während sich der Befragte an die zur Präzisierung benannte Berufsbezeichnung gehalten hat.

Der Befragte gibt als Beruf, zweistufig beantwortet, für Tätigkeit an: „Fließbandarbeit“, und auf Nachfrage nach dem Namen des Berufes: „Staplerfahrer“.

Der Befragte sieht sich als „Fahrzeugführer“ und steigt über den

Code 8: „Maschinen- oder Anlagenbediener, ..., Fahrzeugführer“ ein. Es folgt

Code 83: „Fahrzeugführer“, gefolgt von

Code 8312: „... Rangierer“.

Der Vercoder steigt über das „Fließband“ ein und verortet den Befragten unter Code 8171: „Bediener von automatisierten Montagebändern“.

7 Fazit

Die Frage, die sich stellt, ist die nach der akzeptablen Toleranz bei Abweichungen zwischen Befragtem und Vercoder. Welcher Anteil hiervon lässt sich noch reduzieren und welcher Anteil ist unvermeidlich? Wir werden versuchen, in weiterer Forschung den Anteil der positiven Zuordnungen zu erhöhen, um den unvermeidlichen Anteil der nicht vercodbaren Fälle weiter zu reduzieren.

Die Abweichungen bei der ISCO-Vercodung zwischen Selbstverortung durch die Befragten und Fremdverortung durch die Vercoder scheint sich fast die Waage zu halten. In beiden Fällen betrifft dieses aktuell etwa 15 % der Vercodungen. Wir sind jedoch der Überzeugung, dass sich diese Quote der Abweichungen bei der Selbstverortung noch um einiges reduzieren lässt. Daneben sollten die Befragten nach einem fehlerhaften Einstieg in die Ausgangstabelle eher dazu ermutigt werden, mit einem „weiß nicht“ die nur schwer noch zu bewerkstelligende Selbstverortung abzubrechen. Ein einmal bei „Koch“ verorteter „Bäcker“ ist ohne Klartext nicht mehr zu korrigieren. Zwei Fünftel der fehlerhaften Selbstzuordnungen der Befragten basieren auf einem falschen Einstieg in die Einstiegstabelle, der auf der zweiten bis dritten Stufe der Zuordnung den falschen Weg unterstreicht, da eine korrekte Zuordnung nicht mehr möglich ist. Bei Abbruch würde die verbale An-

gabe des Berufes erforderlich. Die aktuell 10 % Abbrüche würden zusätzlich handvercodet werden. Dieses ergibt eine preisgünstige Vervollständigung der ISCO-Vercodung einer großen Studie.

Bei drei Fünftel der fehlerhaften Zuordnungen (dahinter verbergen sich jedoch nur 8 % der Nennungen) liegen Fehlzuordnungen vor, da der Befragte nicht alle Bedingungen der Definition der ISCO kennt. In diesen Fällen sind zwar die Tätigkeiten in der Regel richtig erfasst, jedoch ist die „Ausbildung“ oder die „Autonomie in der Tätigkeit“ oder eine „Leitungsposition“ bei der Selbstzuordnung nicht gemäß der Definition berücksichtigt worden. Hier lassen sich Fehlzuordnungen in einer Konsistenzprüfung korrigieren.

8 Perspektive

Für den 26. September 2006 ist in Mannheim eine Expertenrunde mit jenen geplant, die entweder nach ISCO-88 vercoden oder die nach ISCO-88 Daten erheben oder die ISCO-88 analysieren. Wir beabsichtigen, in dieser Expertenrunde die Ergebnisse unserer Instrumentenentwicklung vorzustellen und zu diskutieren.

Literatur

- Geis, A. J. & Hoffmeyer-Zlotnik, J. H. P. (2000). Stand der Berufsvercodung. *ZUMA-Nachrichten*, 47, 103-128.
- Geis, A. J. & Hoffmeyer-Zlotnik, J. H. P. (2001). Kompatibilität von ISCO-68, ISCO-88 und KldB-92. *ZUMA-Nachrichten*, 48, 117-138.
- Hoffmeyer-Zlotnik, J. H. P., Hess, D. & Geis, A. J. (2004). Computerunterstützte Vercodung der International Standard Classification of Occupations (ISCO-88). *ZUMA-Nachrichten*, 55, 29-52.
- International Labour Office (Ed.) (1990). *ISCO-88. International Standard Classification of Occupations*. Geneva: ILO.
- Statistisches Bundesamt (Hrsg.) (1992). *Klassifizierung der Berufe – Systematisches und alphabetisches Verzeichnis der Berufsbenennungen*. Stuttgart: Metzler & Poeschel.

Datensätze

- IAB: SGB II-Querschnittserhebung 2006
- infas: ISCO-Pilot 05/2004
- infas: ISCO-Pilot 03/2005
- infas: ISCO-Pilot 08/2005
- infas: ISCO-Pilot 11/2005

Korrespondenzadressen

*PD Dr. Jürgen H.P. Hoffmeyer-Zlotnik
ZUMA
Postfach 12 21 55
68072 Mannheim
E-mail: hoffmeyer-zlotnik@zuma-mannheim.de*

*Alfons J. Geis
ZUMA
Postfach 12 21 55
68072 Mannheim
E-mail: geis@zuma-mannheim.de*

*Doris Hess
infas, Institut für angewandte Sozialwissenschaft GmbH
Friedrich-Wilhelm-Straße 18
53133 Bonn
E-mail: d.hess@infas.de*

1. NUTZERKONFERENZ EVS

„Forschung mit der Einkommens- und Verbrauchsstichprobe (EVS): Analysen zur sozialen Ungleichheit“

19. und 20. Oktober 2006

Hotel Wartburg, Mannheim

Ankündigung

Die Einkommens- und Verbrauchsstichprobe, die vom Statistischen Bundesamt seit 1962 in regelmäßigen 5-Jahres-Abständen durchgeführt wird (und zwar in den Jahren 1962/63, 1969, 1973, 1978, 1983, 1988, 1993, 1998 und 2003) ist das einzige Erhebungsprogramm in Deutschland, das Verbrauch, Ausstattung, Einkommen und Vermögen privater Haushalte in detaillierter Weise mittels laufender Haushaltsbücher erhebt (nähere Informationen hierzu: <http://www.gesis.org/Dauerbeobachtung/GML/Daten/EVS/index.htm> oder http://www.destatis.de/micro/d/micro_c2.htm).

Diese Daten werden in diesem Jahr erstmals als Mikrodaten vollständig für die Forschung in den empirischen Sozial- und Wirtschaftswissenschaften verfügbar sein. Sie stellen ein einzigartiges Potential zur Beschreibung der sozialen Ungleichheit und vor allem zur Untersuchung des sozialen Wandels in Deutschland dar.

Das Ziel der ersten EVS-Nutzerkonferenz ist es, über das Analysepotential der EVS-Mikrodaten zu informieren, neuere, auf der Basis dieser Daten gewonnene Forschungsergebnisse zu diskutieren (insbesondere mit Analysen zum sozialen Wandel von Einkommens- und Verbrauchungleichheiten), wie auch den Erfahrungsaustausch der Nutzer untereinander und mit den statistischen Ämtern als Datenproduzenten zu fördern.

Die Konferenz wendet sich an Forscher, die mit den Scientific oder Public Use Files der EVS bereits arbeiten oder mit diesen zukünftig arbeiten wollen.

Die Konferenz ist eine gemeinsame Veranstaltung von ZUMA und Statistischem Bundesamt.

Konferenzort

Hotel Wartburg, F 4, 4-11, 68159 Mannheim

Übernachtung

Wir schicken Ihnen gerne eine Liste mit Übernachtungsmöglichkeiten zu. Informationen zu Übernachtungsmöglichkeiten erhalten Sie auch bei:

Tourist Information Mannheim
Willy-Brandt-Platz 3
68161 Mannheim
Tel.: 0621 10 10 11
Fax: 0621 2 41 41
E-mail: info@tourist-mannheim.de

Konferenzanmeldung

Die Teilnahme an der Konferenz kostet 65 Euro. Bitte melden Sie sich schriftlich oder per E-mail an, und zwar an folgende Adresse:

ZUMA
Tagungssekretariat
Patricia Lüder
Postfach 12 21 55
D-68072 Mannheim
E-mail: workshop@zuma-mannheim.de

Ansprechpartner bei ZUMA:

Dr. Georg Papastefanou
ZUMA
Tel.: 0621 1246 278
E-mail: papastefanou@zuma-mannheim.de

Ansprechpartner beim Statistischen Bundesamt:

Anette Stuckemeier
Statistisches Bundesamt
Tel.: 01888 644 8800
E-mail: anette.stuckemeier@destatis.de

GESAMMELTE ERFAHRUNG FÜR ALLE VERFÜGBAR

Handbuch der Berufsvercodung im Netz

Das Handbuch gehört zur Grundausrüstung der Codierer/innen für die Berufsvercodung bei ZUMA; es ist Arbeitsgerät und Standard-Nachschlagewerk für alle Fragen zur Codierung nach der Internationalen Standardklassifikation der Berufe (ISCO). Trotz „Standard“ ist es allerdings so kurzlebig wie ein Telefonbuch, denn es wird ständig verändert, ergänzt und korrigiert.

Seitdem die Berufsvercodung bei ZUMA nicht mehr von langjährigen Codierkräften, sondern von studentischen Mitarbeiter/innen mit relativ kurzer Verweildauer durchgeführt wird, war es notwendig geworden, Arbeitsweise, Codierregeln und Konventionen immer stärker zu kodifizieren und vermittelbar zu machen. Was früher vereinzelt auf Handzetteln, Notizen, Listen und Übersichten oder gar nur in den Köpfen vorlag, ist nun systematisch zusammengefasst.

Die Original-Dokumentationen der Berufsklassifikationen umfassen mehrere hundert Seiten; wie aber im konkreten Einzelfall zu entscheiden ist, dazu findet sich oft nichts. Ergänzende Informationen, Hinweise und Entscheidungshilfen, die über das hinausgehen, was in diesen Veröffentlichungen steht, liefert das Handbuch. Damit wird nicht nur die Arbeit der Codierer/innen erleichtert, viel wichtiger ist, dass die Verlässlichkeit der Codierung gewährleistet wird, indem eindeutige Regeln festgelegt werden, nach denen systematisch entschieden werden kann.

Die Dokumentation enthält neben den ISCO-Klassifikationen und ihren ausführlichen Erläuterungen auch Hinweise zur Behandlung von Sonder- und Problemgruppen, wie z.B. Sachbearbeiter und Verwaltungsleute. In der Fallsammlung finden sich immer wiederkehrende Fragen, für die keine allgemeinen Regeln vorliegen, die aber dennoch einheitlich behandelt werden sollen; es wird z.B. festgelegt, wie zu verfahren ist, wenn eine Angabe mehrdeutig ist oder nicht genügend Informationen enthält, um sie einer bestimmten Kategorie zuzuordnen. Auch Randthemen (weil nicht zur eigentlichen Codierung gehörend) werden behandelt; so die Prestige-Generierung, zu der die vollständigen Listen der Werte von SIOPS (Standard International Occupation Prestige-Scale), ISEI (International Socio-Economic Index of Occupational Status) und MPS (Magnitude-Prestigeskala) aufgeführt werden.

Zu finden ist das Handbuch derzeit unter folgendem Link: http://www.gesis.org/Methodenberatung/Textanalyse/Berufs-Branchenklassifikation/Beruf_Branche/hdb_0603_pdf.pdf

In der nächsten Zeit werden in dieser Weise u.a. auch eine kommentierte deutsche Klassifikation der ISCO-88 und das Berufewörterbuch ins Netz gestellt. Anregungen und Anfragen bitte an: Alfons Geis, Telefon 0621/1246-222, geis@zuma-mannheim.de.

CALL FOR PROPOSALS

Themenschwerpunkt „Politische Partizipation, politische Kultur, Sozialkapital“ im ALLBUS 2008

Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) ist ein Serviceangebot der „Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen“ (GESIS) für die sozialwissenschaftliche Forschung und Lehre in Deutschland. Seit 1980 werden alle zwei Jahre Indikatoren zu Einstellungen, Verhaltensweisen und Sozialstruktur der Bevölkerung in der Bundesrepublik Deutschland im ALLBUS erhoben. Die Daten jeder Erhebung stehen unmittelbar nach ihrer Aufbereitung allen interessierten Forscher/innen für Analysen zur Verfügung.

Ziel des ALLBUS ist es, eine inhaltlich und methodisch hochwertige Datenbasis zu schaffen, die den vielfältigen Informations- und Analysebedürfnissen der sozialwissenschaftlichen Profession Rechnung trägt. Für die ALLBUS-Erhebung 2008 ist der Themenschwerpunkt „Politische Partizipation, politische Kultur, Sozialkapital“ ausgewählt worden: Zum einen sollen Teile der Module „Politische Partizipation und Einstellungen zum politischen System“ aus den Jahren 1988 und 1998 repliziert, zum anderen aber auch neue Themen und Fragen zum Zusammenhang von politischer Partizipation, politischer Kultur und Sozialkapital berücksichtigt werden.

Zu Inhalt und Fragenprogramm von ALLBUS 1988 und 1998 siehe: <http://www.gesis.org/Dauerbeobachtung/Allbus/Aktuelles/index.htm>

Mit dem vorliegenden Aufruf möchten wir allen interessierten Wissenschaftler/innen Gelegenheit geben, sich an der Konzeption des Schwerpunktes des ALLBUS 2008 zu beteiligen. Erwünscht sind sowohl Vorschläge für neue Einzelfragen als auch für ganze Fragemodule zu diesem Themenkomplex. Willkommen sind auch kritische Rückmeldungen, Kommentare und Modifikationsvorschläge zu den ALLBUS-Fragen von 1988 und 1998. Da das Fragenprogramm des Schwerpunktes 2008 etwa 25-30 Minuten Befragungszeit umfassen soll, können Neuvorschläge nur in dem Umfang verwirklicht werden, wie Teile des alten Fragenprogramms gestrichen werden.

Die Vorschläge für neue Fragen sollen folgenden Anforderungen entsprechen:

- die Fragen sollten theoriegeleitet sein,
- ihre sozialwissenschaftliche Relevanz im Hinblick auf den Themenschwerpunkt sollte begründet bzw. belegt werden,
- Indikatoren, die sich in allgemeinen nationalen/internationalen Bevölkerungsumfragen bereits bewährt haben, sind willkommen (ggf. Quelle angeben),
- neue Indikatoren sollten mit ausformulierten Frageentwürfen eingereicht werden.

Die Vorschläge sollten nicht länger als acht Seiten sein (ohne Dokumentation bzw. Skizzierung einschlägiger Fragen).

Über die Berücksichtigung der Vorschläge entscheidet der wissenschaftliche Beirat des ALLBUS. Alle Bewerber/innen werden unverzüglich über den Ausgang des Verfahrens informiert. Die ausgewählten Vorschläge werden von der ALLBUS-Abteilung bei ZUMA in Abstimmung mit den Autoren gegebenenfalls weiterentwickelt und getestet.

Für weitere Erläuterungen stehen wir Ihnen gerne zur Verfügung. Wir freuen uns, neue Fragevorschläge und Stellungnahmen zum Fragenprogramm von 2008 bis spätestens 15. September 2006 entgegenzunehmen. Bitte senden Sie Ihre Vorschläge an:

Evi Scholz
Postfach 12 21 55
68072 Mannheim
E-mail: scholz@zuma-mannheim.de
Tel. (0621) 1246-283

BERICHT ÜBER DAS ZWEITE TREFFEN DER ARBEITSGRUPPE MOBILSAMPLE

Tagungsankündigung

Am 21. Februar 2006 fand bei ZUMA das zweite Treffen der Arbeitsgruppe MOBILSAMPLE statt. Diese Arbeitsgruppe, bestehend aus Wissenschaftler/innen in der akademischen und kommerziellen Sozialforschung, ist vor einem Jahr gegründet worden (siehe ZUMA-Nachrichten 56, Mai 2005). Ziel der Arbeitsgruppe ist ein Austausch darüber, wie Mobilfunkanschlüsse in Telefonstichproben integriert werden können.

Beim diesjährigen Treffen war der Teilnehmerkreis gegenüber dem vorjährigen deutlich erweitert. Insgesamt nahmen 14 Sozialwissenschaftler und Statistiker aus 10 Instituten teil. In sieben Präsentationen wurden neueste Entwicklungen auf dem Gebiet der Mobilfunknutzung und Möglichkeiten der Einbeziehung von Mobilfunkteilnehmern in Umfragen vorgestellt. Dabei zeigte sich, dass sich das Problem der exklusiven Handynutzung gegenüber dem Vorjahr weiter verstärkt hat. TNS Infratest ermittelte in seinem Face-to-face-BUS 2005 auf der Basis von ca. 30.000 Interviews einen Anteil von 8,3% der Haushalte, die nur über Mobilfunk erreichbar sind. Zudem werden durch zusätzliche Telefonanbieter (z.B. Voice over Ip) die Probleme der Stichprobengenerierung immer komplexer.

Es wurde übereinstimmend festgestellt, dass das Problem der Telefonstichprobenbildung gegenwärtig von großer Relevanz für die Umfragepraxis ist. Deshalb sollen Forschungsergebnisse zu diesem Thema auf einer Konferenz diskutiert werden, zu der ein größerer Personenkreis Zugang haben wird. Diese Tagung findet am Dienstag, den 21. November 2006 in Mannheim statt. An der Teilnahme Interessierte bitten wir, sich beim Tagungssekretariat des ZUMA (workshop@zuma-mannheim.de) anzumelden.

Korrespondenzadressen

PD Dr. Siegfried Gabler
ZUMA
Postfach 12 21 55
68072 Mannheim
E-mail: gabler@zuma-mannheim.de

Dr. Sabine Häder
ZUMA
Postfach 12 21 55
68072 Mannheim
E-mail: sabine_haeder@zuma-mannheim.de

ZUR ZUKUNFT DER STATISTIKAUSBILDUNG

Forderungen aus der Praxis an die Ausgestaltung der Bachelor- und Master-Studiengänge der Wirtschafts- und Sozialwissenschaften

Als Folge der „Bologna-Erklärung“ zur Schaffung eines einheitlichen europäischen Hochschulraums, die von Wissenschaftsministern aus 45 europäischen Ländern unterzeichnet wurde, werden gegenwärtig an Fachhochschulen und Universitäten Bachelor- und Master-Studiengänge eingeführt. Die gestuften Studiengänge sollen bis 2010 das Regelangebot an deutschen Hochschulen darstellen. Lehre und Studium verändern sich in allen Disziplinen, was auch eine Neuausrichtung der Statistikausbildung zur Folge hat.

Die aktuellen Entwicklungen betrachten wir mit Sorge, weil – nach unserer Beobachtung – der Statistikgrundausbildung in den Curricula für die einschlägigen Bachelor-Studiengänge nicht genügend Aufmerksamkeit geschenkt wird und sie teilweise ganz in die Master-Stufe verlagert wird. Wir befürchten, dass gerade in den berufsqualifizierenden Bachelor-Studiengängen für die Wirtschafts-, Markt- und Sozialforschung wichtige Methodenqualifikationen nicht ausreichend vermittelt werden.

Die Ausbildung in den Bachelor-Studiengängen der Wirtschafts- und Sozialwissenschaften sollte zumindest Grundkenntnisse der Wirtschafts- und Sozialstatistik vermitteln. Durch praxisorientierte Problemstellungen einerseits und durch aktive und kreative Teilnahme an der Lösung der Fragestellungen andererseits sollten die Studierenden auf ihren Übergang in die Arbeitswelt vorbereitet werden. Um dies fruchtbar zu gewährleisten, dürfen die theoretischen Grundlagen nicht vernachlässigt werden, sondern sie sollten zielbezogener, als dies in den Diplomstudiengängen häufig der Fall ist, vermittelt werden. Die bislang vernachlässigten Themen Datenerhebung, Aufbereitung und Auswertung der Daten, Datenschutz und Datenqualität sollten neben den theoretischen Grundlagen (z.B. Messtheorie) bei den Studieninhalten unbedingt berücksichtigt werden. Statistische Methoden sollten unter der Verwendung von echten Daten gelehrt werden, wobei das erworbene theoretische Wissen möglichst in kleinen Projekten oder empirischen Praktika vertieft werden sollte.

In den Master-Studiengängen sollten neue Schwerpunktfächer wie „Empirische Wirtschaftsforschung“ oder „Wirtschaftsstatistik“ geschaffen werden, damit Qualifikationen, die früher in den Wahlfächern Statistik und Ökonometrie erworben wurden, weiter vermittelt werden.

Eine solide Statistikausbildung nimmt im Studium der Wirtschafts- und Sozialwissenschaften eine zentrale Stellung ein, weil in ihr der Umgang mit Informationen, die Basis der Wissensgesellschaft sind, vermittelt wird. Dies setzt allerdings voraus, dass nicht nur formale Methoden gelehrt werden, sondern der gesamte statistische Produktionsprozess von der Operationalisierung einer Fragestellung über die Arbeiten zur Vorbereitung und Durchführung einer Erhebung bis hin zur Datenanalyse und zur Entscheidungsvorbereitung Gegenstand der Ausbildung ist. Ein an den Problemen der Praxis ausgerichtetes Studium verbessert die Berufsaussichten der künftigen Studienabgänger nachhaltig.

Diese Stellungnahme wird unterstützt von (in alphabetischer Reihenfolge):

Prof. Dr. Thomas Bauer
Mitglied des Vorstandes des Rheinisch-Westfälischen Instituts für Wirtschaftsforschung e.V. (RWI)

Prof. Dr. rer. pol. habil. Ulrich Blum
Präsident des Instituts für Wirtschaftsforschung Halle

Thilo Börner
Statistisches Landesamt des Freistaates Sachsen, ver.di-Hauptverwaltung im Statistischen Beirat

Prof. Dr. Dr. h.c. mult. Wolfgang Franz
Präsident des Zentrums für Europäische Wirtschaftsforschung GmbH (ZEW)

Dr. Hans-Peter Glaab
Deutsche Bundesbank

Prof. Dr. Heinz Grohmann
Johann Wolfgang Goethe-Universität Frankfurt (emeritiert)

Johann Hahlen
Präsident des Statistischen Bundesamtes

Hans-Joachim Haß
Bundesverband der Deutschen Industrie e. V.

Prof. Dr. Ullrich Heilemann
Universität Leipzig, Wirtschaftswissenschaftliche Fakultät, Institut für Empirische Wirtschaftsforschung

Prof. Dr. Wim Koesters
Mitglied des Vorstandes des Rheinisch-Westfälischen Instituts für Wirtschaftsforschung e.V. (RWI)

Günter Kromholz
Präsident des Thüringer Landesamtes für Statistik

Prof. Dr. Peter Michael von der Lippe
Universität Duisburg-Essen, Campus Essen, FB 5 Wirtschaftswissenschaften, Fachgebiet Statistik

Dr. Gisela Meister-Scheufelen
Präsidentin des Statistischen Landesamtes Baden-Württemberg

*Prof. Dr. Heiner Meulemann
Institut für Angewandte Sozialforschung
der Universität zu Köln (IfAS). Vorsit-
zender der Arbeitsgemeinschaft Sozial-
wissenschaftlicher Institute e.V. (ASI)*

*Prof. Dr. Peter Meusburger
Geographisches Institut der Universität
Heidelberg. Lehrstuhl für Wirtschafts-
und Sozialgeographie*

*Prof. Dr. Karl Mosler
Universität zu Köln. Seminar für Wirt-
schafts- und Sozialstatistik. Vorsitzender
der Deutschen Statistischen Gesellschaft
(DStatG)*

*Prof. Dr. Gerd Nufer
Professor für BWL mit den Schwerpunk-
ten Marketing und Marktforschung.
Hochschule Reutlingen – Reutlingen
University. School of International Busi-
ness (SIB)*

*Walter Radermacher
Vizepräsident des Statistischen Bundes-
amtes. Stellvertretender Vorsitzender der
Deutschen Statistischen Gesellschaft
(DStatG)*

*Dr. Engelbert Recker
Deutscher Landkreistag*

*Prof. Dr. Ulrich Rendtel
Freie Universität Berlin
Fachbereich Wirtschaftswissenschaft,
Institut für Statistik und Ökonometrie
Vorsitzender des Ausschusses
„Ausbildung und Weiterbildung“ der
Deutschen Statistischen Gesellschaft
(DStatG)*

*Prof. Dr. Ulrike Rockmann
Direktorin des Statistischen Landesamtes
Berlin*

*Manfred Scherschinski
Präsident des Statistischen Landesamtes
Sachsen-Anhalt*

*Prof. Dr. Irene Schneider-Böttcher
Präsidentin des Statistischen Landesam-
tes des Freistaates Sachsen*

*Prof. Dr. Christoph M. Schmidt
Präsident des Rheinisch-Westfälischen
Instituts für Wirtschaftsforschung e.V.
(RWI)*

*Dr. Matthias Schoder
Bereich Wirtschaftspolitik, Mittelstand,
Innovation. Deutscher Industrie- und
Handelskammertag (DIHK)*

*Karl A. Scholz
Bundesverband der Freien Berufe*

*Prof. Dr. Heidemarie Seel
Hochschule für Wirtschaft und Umwelt
Nürtingen-Geislingen*

*Christiane Siegmund
Stellvertretende Geschäftsführerin des
Landesbetriebs für Datenverarbeitung
und Statistik Brandenburg*

*Prof. Dr. Heike Solga
Georg-August-Universität Göttingen.
Sozialwissenschaftliche Fakultät. Institut
für Soziologie*

*Dr. Hans-Jürgen Völz
Volkswirtschaft, Finanzen, Steuern
Bundesvereinigung der Deutschen Ar-
beitgeberverbände*

*Prof. Dr. Gert G. Wagner
Leiter des SOEP und Forschungsdirektor
im DIW Berlin*

*Inhaber des Lehrstuhls „Empirische
Wirtschaftsforschung und Wirtschaftspo-
litik“ an der TU Berlin,
Vorsitzender des Rates für Sozial- und
Wirtschaftsdaten*

*Dr. Peter Weiss
Zentralverband des deutschen Hand-
werks
Wirtschafts- und Umweltpolitik*

*Dr. Robert Weitz
Hauptverband des Deutschen Einzel-
handels*

*Prof. Dr. Horst-Dieter Westerhoff
Bundeskanzleramt*

*Prof. Dr. Klaus F. Zimmermann
Präsident DIW Berlin
Direktor IZA Bonn und
Universität Bonn*

ZUR ZUKUNFT DER STATISTIKAUSBILDUNG

WIESBADEN – Auf unsere Pressemeldung „Zur Zukunft der Statistikausbildung“ vom 9. Februar 2006 haben wir eine breite positive Resonanz erhalten. Insbesondere haben viele weitere Persönlichkeiten erklärt, dieses Anliegen ebenfalls namentlich unterstützen zu wollen.

Diese zusätzlichen Unterstützer finden sich unter dem nachstehenden Text der damaligen Pressemitteilung „Zur Zukunft der Statistikausbildung – Forderungen aus der Praxis an die Ausgestaltung der Bachelor- und Master-Studiengänge der Wirtschafts- und Sozialwissenschaften“.

Als Folge der „Bologna-Erklärung“ zur Schaffung eines einheitlichen europäischen Hochschulraums, die von Wissenschaftsministern aus 45 europäischen Ländern unterzeichnet wurde, werden gegenwärtig an Fachhochschulen und Universitäten Bachelor- und Master-Studiengänge eingeführt. Die gestuften Studiengänge sollen bis 2010 das Regelangebot an deutschen Hochschulen darstellen. Lehre und Studium verändern sich in allen Disziplinen, was auch eine Neuausrichtung der Statistikausbildung zur Folge hat.

Die aktuellen Entwicklungen betrachten wir mit Sorge, weil – nach unserer Beobachtung – der Statistikgrundausbildung in den Curricula für die einschlägigen Bachelor-Studiengänge nicht genügend Aufmerksamkeit geschenkt wird und sie teilweise ganz in die Master-Stufe verlagert wird. Wir befürchten, dass gerade in den berufsqualifizierenden Bachelor-Studiengängen für die Wirtschafts-, Markt- und Sozialforschung wichtige Methodenqualifikationen nicht ausreichend vermittelt werden.

Die Ausbildung in den Bachelor-Studiengängen der Wirtschafts- und Sozialwissenschaften sollte zumindest Grundkenntnisse der Wirtschafts- und Sozialstatistik vermitteln. Durch praxisorientierte Problemstellungen einerseits und durch aktive und kreative Teilnahme an der Lösung der Fragestellungen andererseits sollten die Studierenden auf ihren Übergang in die Arbeitswelt vorbereitet werden. Um dies fruchtbar zu gewährleisten, dürfen die theoretischen Grundlagen nicht vernachlässigt werden, sondern sie sollten zielbezogener, als dies in den Diplomstudiengängen häufig der Fall ist, vermittelt werden. Die bislang vernachlässigten Themen Datenerhebung, Aufbereitung und Auswertung der Daten, Datenschutz und Datenqualität sollten neben den theoretischen Grundlagen (zum Beispiel Messtheorie) bei den Studieninhalten unbedingt berücksichtigt werden. Statistische Methoden sollten unter der Verwendung von echten Daten gelehrt werden, wobei das erworbene theoretische Wissen möglichst in kleinen Projekten oder empirischen Praktika vertieft werden sollte.

In den Master-Studiengängen sollten neue Schwerpunktfächer wie „Empirische Wirtschaftsforschung“ oder „Wirtschaftsstatistik“ geschaffen werden, damit Qualifikationen, die früher in den Wahlfächern Statistik und Ökonometrie erworben wurden, weiter vermittelt werden.

Eine solide Statistikausbildung nimmt im Studium der Wirtschafts- und Sozialwissenschaften eine zentrale Stellung ein, weil in ihr der Umgang mit Informationen, die Basis der Wissensgesellschaft sind, vermittelt wird. Dies setzt allerdings voraus, dass nicht nur formale Methoden gelehrt werden, sondern der gesamte statistische Produktionsprozess von der Operationalisierung einer Fragestellung über die Arbeiten zur Vorbereitung und Durchführung einer Erhebung bis hin zur Datenanalyse und zur Entscheidungsvorbereitung Gegenstand der Ausbildung ist. Ein an den Problemen der Praxis ausgerichtetes Studium verbessert die Berufsaussichten der künftigen Studienabgänger nachhaltig.

Zusätzliche Unterstützer:

Prof. Ph. D. Jutta Allmendinger
Direktorin des Instituts für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit

Dr. Peter Bauer
Präsident des Bayerischen Landesamtes für Statistik und Datenverarbeitung

Stefan Bender
Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit.
Leiter des Forschungsdatenzentrums der Bundesagentur für Arbeit im IAB

Prof. Axel Börsch-Supan, Ph.D.
Mannheim Research Institute for the Economics of Aging (MEA), Universität Mannheim

Eckart Hohmann
Präsident des Hessischen Statistischen Landesamtes

Prof. Dr. Reinhard Hujer
Johann Wolfgang Goethe-Universität Frankfurt am Main

Prof. Dr. Michael Hüther
Direktor und Mitglied des Präsidiums
Institut der deutschen Wirtschaft Köln

Prof. Dr. Gerhard Kleinhenz
Universität Passau. Lehrstuhl für Volkswirtschaftslehre mit Schwerpunkt Wirtschafts- und Sozialpolitik

Prof. Dr. Dr. h.c. Hans-Jürgen Krupp
Darmstadt

Prof. Dr. Steffen-Matthias Kühnel
Georg-August-Universität Göttingen,
Methodenzentrum Sozialwissenschaften

Prof. Dr. Peter Ph. Mohler
Direktor des Zentrums für Umfragen, Methoden und Analysen (ZUMA), Mannheim

Prof. Dr. Dr. Ulrich Mueller
Institut für Medizinische Soziologie und Sozialmedizin. Klinikum Universität Marburg

Prof. Dr. Hans Rattinger
Universität Bamberg. Lehrstuhl für Politikwissenschaft II

*Uwe G. Rehfeld
Deutsche Rentenversicherung Bund.
Leiter der Abteilung Volkswirtschaft,
Statistik und Empirische Analysen zur
Alterssicherung und Altersvorsorge*

*Prof. Regina T. Riphahn, Ph.D.
Universität Erlangen-Nürnberg*

*Dr. Sibylle Schmerbach
Humboldt-Universität zu Berlin. Wirt-
schaftswissenschaftliche Fakultät. Pro-
dekanin für Studium und Lehre*

*Prof. Dr. Peter Schmidt
Hochschule Bremen. Fachbereich Wirt-
schaft*

*Prof. Dr. Reiner Stäglin
Freie Universität Berlin Fachbereich
Wirtschaftswissenschaft, Institut für
Statistik und Ökonometrie. Stellvertre-
tender Vorsitzender der Deutschen Sta-
tistischen Gesellschaft*

*Prof. Dr. Thomas Straubhaar
Direktor des Hamburgischen WeltWirt-
schaftsinstituts (HWWI). Präsident des
Hamburgischen Welt-Wirtschafts-Archiv
(HWWA). Universitätsprofessor für
„Volkswirtschaftslehre, insbesondere
Wirtschaftspolitik“ der Universität
Hamburg*

*Prof. Dr. Götz Uebe
Helmut-Schmidt-Universität. Universität
der Bundeswehr Hamburg*

*Prof. Dr. Gerhard Wagenhals
Universität Hohenheim. Lehrstuhl Statis-
tik und Ökonometrie. Institut für Volkswirt-
schaftslehre*

*Prof. Dr. Joachim Wagner
Universität Lüneburg. Empirische Wirt-
schaftsforschung. Institut für Volkswirt-
schaftslehre*

*Dr. Eckhardt Wohlers
HWWA-Institut für Wirtschaftsfors-
chung-Hamburg e.V. Leiter der Abtei-
lung Internationale Makroökonomie*

*Dr. Sylvia Zühlke
Landesamt für Datenverarbeitung und
Statistik Nordrhein-Westfalen.
Leiterin der Forschungsdatenzentren der
Statistischen Landesämter*

*Markus Zwick
Statistisches Bundesamt
Leiter des Forschungsdatenzentrums*

Weitere Auskünfte gibt:

*Dr. Manfred Ehling
Telefon: (0611) 75-2903
E-Mail: manfred.ehling@destatis.de*

BUCHBESPRECHUNGEN

Kognitive Interviews haben sich seit ca. 10 Jahren als Standard-Tool zur Evaluation von Survey-Fragen und Survey-Fragebogen etabliert. Erfreulicherweise liegt jetzt auch ein Buch vor, das sich – als erstes seiner Art – ausschließlich dieser Technik widmet. Der Autor Gordon Willis ist ein ausgewiesener Experte auf diesem Gebiet. Er arbeitet als Cognitive Psychologist am National Cancer Institute in den USA und hat bereits zahlreiche Artikel und Beiträge zu diesem Thema veröffentlicht.

GORDON B. WILLIS

Cognitive Interviewing. A Tool for Improving Questionnaire Design

Thousand Oaks: Sage Publications, 2004

ISBN 0-7619-2804-9, 335 Seiten, \$ 35,95

Das Buch ist in vier Teile gegliedert, die wiederum jeweils aus 3 - 4 Kapiteln bestehen. Jedes Kapitel wird von einer Zusammenfassung abgeschlossen.

Nach einer Definition kognitiver Interviews wird in Teil 1 eine Einführung in das Thema gegeben und dabei anschaulich erläutert, wie kognitive Interviews in dem weiten Kontext der Fragebogenkonstruktion verankert sind. Sehr hilfreich ist hier z.B. eine Übersicht mit Definitionen von allgemeinen Merkmalen bzw. Eigenschaften kognitiver Test-Prozeduren, z.B. wird hier vom Autor klargestellt, dass kognitive Interviews zwar keine „beweiskräftigen“ Daten liefern, jedoch ein äußerst nützliches informationsgenerierendes Werkzeug zum Testen von Survey-Fragen sind. Der Autor führt drei Argumente an, weshalb Fragen grundsätzlich getestet werden müssen. Im letzten Kapitel dieses Teils werden kognitive Techniken, wie z.B. die Technik des lauten Denkens oder verschiedene Nachfragetechniken ausführlich beschrieben und ihre Vor- und Nachteile diskutiert. Abschließend führt der Autor eine Reihe von Argumenten auf, weshalb Fragen eigentlich getestet werden müssen.

Teil 2 trägt den Titel „The Intricacies of Verbal Probing“, was so viel heißt wie „Die Kniffligkeit des Nachfragens“ und befasst sich mit der Technik sowie den vielfältigen Problemen des Nachfragens (Verbal Probing). Das bis heute nicht befriedigend geklärte Problem, nämlich welche Fragen hinterfragt werden sollen und welche Nachfragen gestellt werden sollen, wird vom Autor umfassend und verständlich diskutiert. Für den Leser sind nicht nur die ausführlich beschriebenen sieben Beispiele aus der Praxis, sondern auch

das vom Autor entwickelte Checksystem, das Hilfestellung bei der Identifizierung potentieller Frageprobleme geben kann sowie eine Auflistung von Regeln zur Anwendung von „Conditional Probes“ äußerst hilfreich. Abschließend macht der Autor Vorschläge zur Entwicklung kognitiver Standard-Nachfragen und gibt konkrete Ratschläge, wie man Fallstricke beim Nachfragen vermeidet.

Teil 3 beinhaltet sehr praxisorientierte Informationen im Hinblick auf die Vorbereitung und Durchführung kognitiver Interviews. Das erste Kapitel befasst sich sehr ausführlich mit verhaltensbedingten Merkmalen und inhaltlichen Kompetenzen, die ein kognitiver Interviewer besitzen muss. Weiterhin wird ausgeführt, wie kognitive Interviewer sich von „Feld-Interviewern“ unterscheiden und im Hinblick auf ihre Aufgaben zu schulen sind. Die folgenden zwei Kapitel beschreiben sehr detailliert die für die Vorbereitung und Durchführung kognitiver Interviews notwendigen Arbeitsschritte. Dabei ist die große praktische Erfahrung des Autors spürbar, vor allem dann, wenn er auf Details einer kognitiven Testphase eingeht. Beispielsweise werden hier Fragen beantwortet wie: Nach welchen Merkmalen und auf welche Art und Weise werden Zielpersonen rekrutiert? Wie viele Fragen können in einem kognitiven Interview getestet werden und wie lange kann ein solches maximal dauern? Wie viel Interviews müssen durchgeführt werden, um verlässliche Ergebnissen zu erhalten? Welche Aspekte sind im Hinblick auf den Datenschutz zu beachten? Wie muss ein Raum aussehen, in dem ein kognitives Interview stattfindet? Der Autor gibt die Antworten auf der Grundlage seiner eigenen umfangreichen Erfahrungen, räumt aber ein, dass diese nicht generell verbindlich sein müssen. Im letzten Kapitel werden Dokumentation und Analyse der Ergebnisse kognitiver Interviews behandelt, wobei es zu beachten gilt, dass die Ergebnisse eines kognitiven Tests nicht automatisch zu einer verbesserten Frageversion führen, sondern ein erhebliches Maß an Entscheidungen und Interpretation erforderlich ist, um letztendlich zu einem verbesserten Fragendesign zu kommen. In dem Buch werden zwei Analyseverfahren vorgestellt: ein eher informell-qualitatives, bei dem die Ergebnisse beschrieben werden und ein strukturiert-quantitatives, bei dem die Ergebnisse mit Hilfe von Kategorienschemata codiert werden. Beide Vorgehensweisen werden beispielhaft an den in diesem Buch vorrangig behandelten kognitiven Techniken des lauten Denkens (Think-Aloud) und des Nachfragens demonstriert und ihre Vor- und Nachteile beschrieben. Zwei Übungen beenden diesen Teil. Der Leser wird dabei aufgefordert, auf Grund der Reaktionen von Testpersonen das Problem der vorgegebenen Frage zu diagnostizieren und eine Lösung bzw. Verbesserung vorzuschlagen. Das Ergebnis kann anschließend mit einer vom Autor vorgeschlagenen Lösung verglichen werden.

In Teil 4 wird zunächst beschrieben, wie kognitive Interviews in speziellen Anwendungsbereichen eingesetzt werden können. Der Leser erfährt hier z.B., wie kognitive Interviews an unterschiedliche Erhebungsmodi (Face-to-Face, Telefon, Selbstausfüller, computerge-

stützt) angepasst werden können und erhält u.a. wichtige und interessante Hinweise darauf, was besonders beim Testen von heiklen Fragen oder Fragen für Unternehmensbefragungen zu beachten ist. Darüber hinaus geht Gordon Willis auch auf Besonderheiten ein, die beim Interviewen von sehr jungen und sehr alten Testpersonen berücksichtigt werden müssen. Im nachfolgenden Kapitel, das bezeichnenderweise den Untertitel „What is truth?“ trägt, wird die Leistungsfähigkeit kognitiver Techniken kritisch beleuchtet, wobei der zentralen Frage nachgegangen wird „How do we know that the cognitive approach to testing survey questions really works?“ Dies geschieht einerseits aus theoretischer Sicht, andererseits wird unter Zuhilfenahme mehrerer Evaluationsmodelle versucht zu klären, inwieweit die Ergebnisse empirischer Evaluationsstudien zur Beantwortung der Frage beitragen können. Das Kapitel wird abgeschlossen von einem Vergleich kognitiver Interviews mit anderen Pretestmethoden, wie z.B. Expert Reviews, Focus Groups, Behaviour Coding und Reinterviews. Im letzten Kapitel gibt der Autor zunächst 12 ausführlich beschriebene Empfehlungen zur praktischen Durchführung kognitiver Interviews und äußert sich abschließend zu zukünftigen Entwicklungen.

Die beiden Anhänge sind für jeden Praktiker eine wahre Fundgrube, und zwar sowohl für den Anfänger auf dem Gebiet des kognitiven Interviewens, als auch für den Experten. Der erste Anhang enthält das Beispiel eines kompletten vorstrukturierten Fragebogens für ein kognitives Interview - mit den zu testenden Fragen, vorgegebenen Nachfragen und Instruktionen für den Interviewer (insgesamt 13 Seiten). Im zweiten Anhang werden aus Original-Testberichten die Ergebnisse zu einzelnen Fragen zitiert, die zusätzlich vom Autor kritisch kommentiert werden.

Alles in allem ist das Buch das gelungene Ergebnis jahrelanger Erfahrungen mit der Qualitätsüberprüfung von Survey-Fragen. Gordon Willis beherrscht nicht nur das Handwerkzeug kognitiver Interviews perfekt, er versteht es auch ausgezeichnet, diese Testmethode im größeren Kontext der Fragebogenentwicklung zu verorten. Der Autor stützt sich aber nicht nur auf sein profundes Wissen, sondern verweist auch häufig auf neueste Forschungsarbeiten und Literatur zum Thema. Das Buch ist gut gegliedert und die Zusammenfassungen am Ende jedes Kapitels sind sehr hilfreich, außerdem sind die in den Teilen 1, 2 und 3 enthaltenen Übungen hervorragende Arbeitshilfen.

Das Buch – das erste Lehrbuch für die Durchführung kognitiver Interviews überhaupt – ist ein absolutes „Muss“ für alle diejenigen, die sich mit der Überprüfung der Qualität von Fragen beschäftigen – unabhängig davon, ob es sich dabei um Anfänger oder Experten handelt. Wir können das Buch uneingeschränkt empfehlen.

PETER PRÜFER & MARGRIT REXROTH

Die im Herbst 2005 erschienene Studie von Nina Baur greift die Frage nach den methodologischen Konsequenzen der Zeitlichkeit sozialen Handelns auf. Sie leistet einen wichtigen Beitrag, indem sie den Versuch unternimmt, dieses Forschungsfeld systematisch und mit empirischen Beispielanalysen zu beleuchten.

NINA BAUR

Verlaufsmusteranalyse.

Methodologische Konsequenzen der Zeitlichkeit sozialen Handelns

Wiesbaden: VS Verlag für Sozialwissenschaften, 2005

ISBN 3-531-14727-7, 367 Seiten, € 39,90

Der Band gliedert sich in drei Teile. Im ersten Teil wird der Frage nachgegangen, warum sich die Soziologie bislang wenig um theoretische Konzepte und Methoden zur Erfassung von Verlaufsmustern bemüht hat. Sie findet den Grund in der disziplinären Arbeitsteilung zwischen Soziologie und Geschichtswissenschaft, die sich zwischen 1920 und 1950 aufgrund historischer Zufälle und wissenschaftspolitischer Entscheidungen zu getrennten Disziplinen entwickelt haben. Während sich Historiker von jeher intensiv mit verschiedenen Aspekten der Zeitlichkeit sozialen Handelns befasst haben, beschäftigt sich die Soziologie erst Ende der 70er Jahre wieder mit der Zeitlichkeit sozialen Handelns. Aus der detaillierten wissenschaftshistorischen Untersuchung zieht die Autorin die Schlussfolgerung, dass die soziologische Betrachtung zeitlicher Verläufe mit den Erkenntnissen der Geschichtswissenschaft zusammengeführt werden sollten.

Darauf aufbauend ist der anschließende Teil des Buches darauf gerichtet, einen kategorialen Bezugsrahmen für die sozialwissenschaftliche Analyse von Verlaufsmustern zu entwickeln. Wesentliche Elemente sind eine Ereignismatrix aus Ereignissen und Ereignisketten, Formen von Verlaufsmustern, zeitliche Schichten in Form unterschiedlicher Dauer und die Verschränkung von Verlaufsmustern unterschiedlicher Handlungsbereiche.

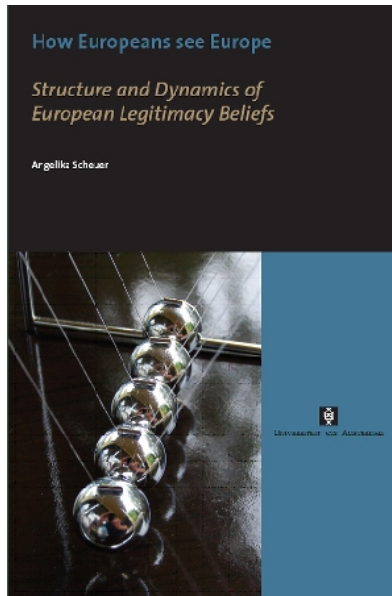
Mit Hilfe dieses Bezugsrahmens untersucht die Autorin dann im dritten Teil mehrere in der quantitativen Sozialforschung entwickelte Methoden wie Kohorten-, Zeitreihen-, Ereignis- und Sequenzanalyse daraufhin, inwiefern sie zur Erfassung von Verlaufsmustern geeignet sind. Diese Verfahren werden in ihren Grundzügen skizziert, ihr analytisches Potential wird anhand einer einschlägigen empirischen Untersuchung beispielhaft erläutert.

Nina Baur beschränkt ihre Betrachtung jedoch nicht nur auf quantitative Ansätze, sondern lenkt das Augenmerk auch auf Beispiele der qualitativen Sozialforschung. Ansätze der grounded theory, der biographischen Methode, von Fallstudien und historisch-vergleichenden Arbeiten werden ebenfalls anhand veröffentlichter Analysen daraufhin untersucht, inwiefern sie Wege zur Umsetzung der Perspektive von Verlaufsmustern darstellen.

Die mit diesem Band vorgelegte gründliche und umfassende Beschäftigung mit der Methodologie von verlaufsorientierten Untersuchungsansätzen führt verschiedene statistische und methodische Instrumente zusammen. Mit einem eigenen, allgemeinen Kategoriensystem ermöglicht sie es, sonst disparat und abgeschottet von einander existierende Ansätze vergleichend zu betrachten. Dies ist ein anerkennenswerter Anspruch, dessen Umsetzung in diesem Band dazu beitragen kann, der Zersplitterung als Konsequenz der methodologischen Spezialisierung entgegenzuarbeiten. Dem Buch von Nina Baur ist eine breite disziplinübergreifende Resonanz zu wünschen.

GEORGIOS PAPASTEFANOU

HOW EUROPEANS SEE EUROPE. STRUCTURE AND DYNAMICS OF EUROPEAN LEGITIMACY BELIEFS



Seit die Europäische Union nicht nur die wirtschaftliche Integration, sondern auch die politische Vereinigung zum Ziel hat, haben Forscher wiederholt auf den erhöhten Bedarf an politischer Legitimität – im Sinne der Unterstützung der Bürger für die europäische Integration und das politische System der EU – hingewiesen. Jedoch bot die Forschung bisher keine überzeugende Handhabe für die Modellierung der dynamischen Prozesse, die sie theoretisch postuliert. Zwei methodische Probleme behinderten den empirischen Zugang zur Entwicklung der europäischen Legitimität: die Messung europäischer Einstellungen und die Modellierung dynamischer Prozesse. Der in dieser Doktorarbeit entwickelte Ansatz erlaubt es, die europäischen Legitimitätseinstellungen adäquat zu messen, die Legitimitätsprozesse zu modellieren und somit alternative Legitimitätshypothesen zu testen.

Die Erfassung von Legitimitätseinstellungen basierte zumeist auf Ad-hoc-Messung, d.h. der Zuweisung von Konzepten zu Surveyfragen aufgrund von Fragetexten. Diese Zuweisung variierte zwischen Studien und machte die Ergebnisse unvergleichbar. Auch die Äquivalenz von Konzepten und Indikatoren zwischen den EU-Ländern blieb eine offene Frage. Diese Analyse geht von der Annahme aus, dass die Einstellungen zur EU ein System bilden, und mündet in einer empirisch fundierten Darstellung, wie die Bürger ihre Einstellungen zur EU mental organisieren. Dazu wird ein mehrstufiges Verfahren der Strukturanalyse entwickelt und auf eine Umfrage der Europawahlstudie von 1994 mit allen (damals) 12 Mitgliedstaaten angewendet.

Das Ergebnis besteht im Modell eines differenzierten und klar strukturierten Einstellungssystems, das in allen 12 EU-Ländern anwendbar ist. Die Struktur wird von Einstellungsmodi – kognitiv, affektiv und evaluativ – und der Unterscheidung zwischen zwei Objekten – wirtschaftliche Integration und politische Vereinigung – bestimmt. Vier Einstellungsdimensionen lassen sich nachweisen: Kognition, Evaluation, Affekt für wirtschaftliche Integration und Affekt für politische Vereinigung. Anhand dieses Einstellungsmodells kann die Bedeutung häufig verwendeter Indikatoren geklärt und eine valide und komparative Messung relevanter Konzepte entwickelt werden.

Die Modellierung der dynamischen Prozesse, die in den Legitimitätshypothesen formuliert werden, mit Querschnittsdaten ist das zweite Problem der Legitimitätsforschung. Während die Integrationstheorie annimmt, dass die Bürger die EU legitimieren, weil diese von den Eliten als erstrebenswert präsentiert wird, erwartet die Nutzentheorie, dass Legitimität aus positiven Leistungen der EU entsteht. Der einflussreiche Ansatz von David Easton (1964) integriert beide Prozesse. Das länderübergreifende Einstellungsmodell macht die EU-12-Länder zu einem Labor für komparative Hypothesentests. Dazu wird – aufbauend auf Erkenntnisse der assoziativen Netzwerkforschung – ein Ansatz entwickelt, der aus gegenwärtigen Strukturen Erkenntnisse über die Dynamik vergangener Prozesse ableitet. Anhand von Strukturgleichungsmodellen werden zunächst die externen Faktoren ermittelt, die auf die europäischen Einstellungsdimensionen einwirken. Danach werden die dynamischen Prozesse innerhalb des Einstellungssystems durch iterative Anwendung von nicht-rekursiven Modellen identifiziert. Vier Typen der Legitimitätsentwicklung – oder Legitimationsstile – werden ermittelt. Drei davon zeichnen sich durch die Dominanz der diffusen Unterstützung für die politische Vereinigung aus. Beim vierten Legitimationsstil ist die spezifische Unterstützung für die wirtschaftliche Integration prägend für die Haltung gegenüber der EU. Die empirischen Ergebnisse stützen die Hypothesen der Integrationstheorie und des Easton'schen Ansatzes, wohingegen ein strikt utilitaristischer Ansatz widerlegt wird.

**How Europeans see Europe.
Structure and Dynamics of European Legitimacy Beliefs**

Angelika Schauer

Amsterdam University Press, Amsterdam 2005

ISBN: 90 5629 408 3, 228 Seiten, 37,50 €

ZUMA-NACHRICHTEN SPEZIAL

Band 11:

Methodological Aspects in Cross-National Research

Jürgen H.P. Hoffmeyer-Zlotnik & Janet A. Harkness (Eds.)

Mannheim: ZUMA 2005, 305 Seiten, ISBN 3-924220-29-8, 13 Euro

The idea for this volume was born during the Sixth International Conference on Social Science Methodology in Amsterdam in August 2004, organized by the International Sociological Association Research Committee 33 on Logic and Methodology. Most of the contributions in this volume are proceeding papers from the Amsterdam conference. The contributions in this volume are organized in four parts. The first part deals with designing and implementing cross-cultural surveys. The second part consists of three papers that deal with different issues of comparability or “equivalence”. The third part of the volume brings together papers on with harmonizing socio-demographic information in different types of surveys. The last section of the volume contains papers that discuss individual socio-demographic variables in cross-national perspective.

Band 12:

Conducting Cross-National and Cross-Cultural Surveys. Papers from the 2005 Meeting of the International Workshop on Comparative Survey Design and Implementation (CSDI)

Janet A. Harkness (ed.)

Mannheim: ZUMA 2006, 123 Seiten, ISBN 3-924220-31-X, 13 Euro

The papers in this volume stem from the third annual meeting of the International Workshop on Comparative Survey Design and Implementation (CSDI). Initiated in 2002, the Workshop developed out of cross-cultural symposia held at ZUMA throughout the nineteen nineties. One of CSDI’s primary goals is to promote research into methodological issues of particular and sometimes unique salience for cross-cultural and cross-national survey research. For more information visit the CSDI website (www.csd-workshop.org).

The seven papers are good illustrations of the broad spectrum of research fields in which CSDI researchers are engaged. The volume begins and ends with two framework papers, the first discussing what makes cross-national research special, the last on where we begin to draw boundaries between entities to be compared in “comparative” research. The five remaining papers discuss (in order of the volume): the rich information available from the multinational European Social Survey on data collection; socio-demographic measure-

ment and comparability in the cross-national context, again with reference to the European Social Survey; cognitive pre-testing of translated questionnaires; communicative issues across cultures in telephone interviews; and preliminary work on guidelines on using interpreters underway at the U.S. Census Bureau. The last-mentioned papers reflect research concerns in U.S. cross-cultural context.

ZUMA-ARBEITSBERICHTE

Die „ZUMA-Arbeitsberichte“ informieren über das gesamte Spektrum der Arbeit bei ZUMA und enthalten Beiträge, die zur Veröffentlichung in Fachzeitschriften vorgesehen sind. Ab 1996 (und vereinzelt zuvor) sind die ZUMA-Arbeitsberichte als PDF-Datei abrufbar (http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/); ebenfalls ab 1996 informieren Abstracts über den Inhalt der einzelnen Arbeitsberichte.

Arbeitsberichte, die vor 1996 veröffentlicht wurden, können ebenfalls unter http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/ bestellt werden; bei einer maximalen Anzahl von vier solcher Arbeitsberichte erfolgt die Auslieferung kostenlos, bei Bestellung von mehr als vier Heften werden pro Heft 3 € in Rechnung gestellt.

Rolf Porst & Cornelia Jers:

Die ALLBUS-„Gastarbeiter-Frage“. Zur Geschichte eines Standard-Instruments in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS). ZUMA-Arbeitsbericht 2005/04.

Die „Gastarbeiter-Fragen“ sind mittlerweile zu einem Standardinstrument nicht nur des ALLBUS, sondern auch für andere Umfragen geworden, die sich mit Einstellungen zu ausländischen MitbürgerInnen befassen. In dem Arbeitsbericht werden Entstehung, Entwicklung und Ergebnisse dieser Frage beschrieben. Man bekommt unter anderem auch Hinweise darauf, dass und wie man mit Hilfe eines split half-Verfahrens einen optimalen Umstieg von einer antiquierten hin zu einer moderneren Frageformulierung schafft, ohne den Zeitreihencharakter einer Frage zu beschädigen. Darüber hinaus zeigt die Entwicklung der „Gastarbeiter-Frage“, dass die Notwendigkeit und Sinnhaftigkeit eines solchen Umstiegs bereits durch kognitive Pretestverfahren aufgezeigt und nahegelegt werden kann.

Andrea Lengerer, Jeanette Bohr & Andrea Janßen:

Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen. ZUMA-Arbeitsbericht 2005/05.

Als Haushaltsstichprobe enthält der Mikrozensus Informationen über die Größe und Zusammensetzung von Haushalten sowie über die Beziehungen der Haushaltsmitglieder untereinander. Damit lassen sich sowohl die Haushaltsstrukturen als auch die Familien-

formen der Bevölkerung untersuchen. Die in der amtlichen Statistik verwendeten Konzepte und Definitionen sowie die darauf aufbauenden Typologien sind hierzu von grundlegender Bedeutung. Diese werden in dem vorliegenden Bericht dargestellt und in ihrer konkreten Umsetzung erläutert.

Neben Haushalten und Familien werden seit 1996 auch Lebensformen als soziale Einheiten in den Daten abgegrenzt. Im Mittelpunkt des neuen Konzepts steht die Berücksichtigung unverheiratet zusammenlebender Paare und die Familientypologie ist um eine Typologie der Lebensformen ergänzt. Das familienwissenschaftliche Analysepotenzial des Mikrozensus hat sich dadurch beträchtlich erweitert, gleichwohl bleiben Grenzen der Abbildung familialer und partnerschaftlicher Lebensformen bestehen.

Matthias Fleck & Georgios Papastefanou:

Einkommens- und Verbrauchsstichprobe 1998 – Design und Methodik sowie Veränderungen gegenüber den Vorgängererhebungen. ZUMA–Arbeitsbericht 2006/01.

Mit der achten Auflage des Jahres 1998 wurde in der Einkommens- und Verbrauchsstichprobe (EVS) eine Reihe weitreichender inhaltlicher und methodischer Veränderungen realisiert. Die Quotierungsmerkmale wurden leicht verändert, ein neues Hochrechnungsverfahren eingesetzt, Erhebungsteile umstrukturiert, die Anschreibephase in den Haushaltsbüchern wurde verkürzt, der Merkmalskatalog wurde überarbeitet und ein neues, international vergleichbares, Klassifikationsschema implementiert. Die Diskussion dieser Modifikationen und die Beurteilung ihrer Konsequenzen für die Datenqualität und die Vergleichbarkeit mit früheren Einkommens- und Verbrauchsstichproben sind Gegenstand dieses Arbeitsberichts. Darüber hinaus wird der für den Zeitvergleich erforderliche Umsteigeschlüssel dargestellt und erläutert.

ZUMA-METHODENBERICHTE

In der ZUMA-Reihe „Methodenberichte“ werden Beiträge publiziert, die wichtige Aspekte der Arbeit von ZUMA dokumentieren, aber in der Regel nicht für eine Veröffentlichung in einer Fachzeitschrift geeignet sind, wie z.B. Dokumentationen, Bibliographien, Literaturberichte und Datenbeschreibungen. Die Berichte werden in der Regel nicht verschickt, sondern stehen als PDF-Dateien im Internet zur Verfügung:

(www.gesis.org/publikationen/berichte/zuma_methodenberichte/)

Gedruckte Einzelhefte können bei den jeweiligen Autoren angefordert werden.

Evi Scholz, Janet A. Harkness & Timo Faäß

ALLBUS-Bibliogr ISSP Study Monitoring 2003 Report to the ISSP General Assembly on monitoring work undertaken for the ISSP by ZUMA, Germany. ZUMA-Methodenbericht 2005/08

Evi Scholz & Janet A. Harkness

ISSP 2003 Germany National Identity II ZUMA Report on the German Study. ZUMA-Methodenbericht 2005/09

Evi Scholz & Janet A. Harkness

ISSP 2004 Germany Citizenship I ZUMA Report on the German Study. ZUMA-Methodenbericht 2005/10

Yvonne Lechert & Paul Lüttinger

Mikrozensus Scientific Use File 1973. Dokumentation und Datenaufbereitung. ZUMA-Methodenbericht 2006/01

Yvonne Lechert & Paul Lüttinger

Mikrozensus Scientific Use File 1976. Dokumentation und Datenaufbereitung. ZUMA-Methodenbericht 2006/02

Yvonne Lechert & Paul Lüttinger

Mikrozensus Scientific Use File 1982. Dokumentation und Datenaufbereitung. ZUMA-Methodenbericht 2006/03

Yvonne Lechert & Paul Lüttinger

Mikrozensus Scientific Use File 1973. Dokumentation und Datenaufbereitung. ZUMA-Methodenbericht 2006/04

Andrea Lengerer & Mara Boehle

Rekonstruktion von Bandsatzerweiterungen zu Haushalt, Familie und Lebensformen im Mikrozensus ZUMA-Methodenbericht 2006/05

Alexander Haarmann, Evi Scholz, Martina Wasmer Michael Blohm & Janet A. Harkness

Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 2004 ZUMA-Methodenbericht 2006/06

Julia H. Schroedter & Kathrin Leim

Mikrozensus Scientific Use File 2004: Vergleich mit dem vorhergehenden Erhebungszeitpunkt ZUMA-Methodenbericht 2006/07

ZUMA-WORKSHOPS

Zweites Halbjahr 2006

Thema	Datum
Grundlagen der Fragebogenentwicklung	12. – 13. September
Das Mikrozensus-Panel 1996 – 1999 als Scientific Use File: Eine Einführung	14. September
Simulation for the Social Sciences	18. – 22. September
Einführung in die Analyse linearer Strukturgleichungsmodelle mit LISREL	27. – 29. September
Ausländische Bevölkerung im Mikrozensus	5. Oktober
Methoden zur Analyse latenter Einstellungsstrukturen	6. Oktober
Stichproben in Theorie und Praxis	11. Oktober
Qualitative Inhaltsanalyse	30. – 31. Oktober
Grounded Theory	7. – 8. November
Qualitative Interviews – Theorie und Praxis	10. – 11. November
Einführung in die computerunterstützte Inhaltsanalyse (cui)	14. – 15. November
Testung sozialwissenschaftlicher Erhebungsinstrumente mit kognitionspsychologischen und psychometrischen Methoden	16. – 17. November
Grundlagen der Fragebogenentwicklung	28. – 29. November
Praktische Durchführung der Telefonbefragungen	1. Dezember
Online-Umfragen	11. – 12. Dezember

Detaillierte Hinweise zu den oben aufgeführten Veranstaltungen erhalten Sie unter:

<http://www.gesis.org/Veranstaltungen/ZUMA/Workshops/index.htm>

oder beim ZUMA-Tagungssekretariat (workshop@zuma-mannheim.de, Tel.: 0621/1246221).

NACHRICHTEN AUS DEM INSTITUT

In den letzten Monaten wurden Mitarbeiterinnen und Mitarbeiter von ZUMA erneut mit besonderen Anerkennungen und Ehrungen für ihre Arbeit gewürdigt.

In erster Linie ist hier die Verleihung des RENÉ-DESCARTES-PREIS AN DEN EUROPEAN SOCIAL SURVEY (ESS) für seine Einführung radikaler Neuerungen bei länderübergreifenden Erhebungen zu nennen. Das Team um ROGER JOWELL von der City University London (Großbritannien) und PETER MOHLER von ZUMA erhielt für eines der großen, europäischen Umfrageprojekte zur Darstellung und Erklärung von sozialem, politischem und einstellungsbezogenem Wandel in und zwischen den europäischen Staaten 200.000 Euro. Das Besondere am ESS sind jedoch nicht die Inhalte, sondern die strengen methodischen Mindeststandards für interkulturell vergleichende Studien, und damit einhergehend seine Benchmark-Funktion für andere europäischen Studien. Zudem bietet der ESS eine einzigartige Plattform für Akademiker und Politiker auch ohne tiefer gehende statistische Kenntnisse für Informationen und Daten als Entscheidungsgrundlagen zu den wichtigen gesellschaftlichen Themenbereichen.

Eine besondere Anerkennung erfuhr CAROLINE KRAMER (Abteilung Soziale Indikatoren, zurzeit beurlaubt), die für ihre Habilitation „Zeit für Mobilität. Räumliche Disparitäten der individuellen Zeitverwendung für Mobilität in Deutschland“ den mit 15.000 Euro dotierten Dr. Meyer-Struckmann-Wissenschaftspreis 2005 der Brandenburgischen Technischen Universität Cottbus entgegennehmen konnte. Für ihre Diplomarbeit „Binationale Ehen in Deutschland – Der Einfluss der strukturellen auf die soziale Integration“ wurde JULIA SCHROEDTER (German Microdata Lab) mit dem Gerhard-Fürst-Preis 2005 geehrt.

Anerkennung erfahren hat auch CHRISTOF WOLF, als wissenschaftlicher Leiter für den Bereich Dauerbeobachtung zuständig: Das von ihm beantragte Projekt „Sozialer und ökonomischer Wandel in (West-) Deutschland“ wird im Rahmen des Paktes für Forschung und Innovation von der Leibniz-Gemeinschaft drei Jahre lang gefördert.

ZUMA gratuliert ANGELIKA SCHEUER (Abteilung Soziale Indikatoren), die sich im Dezember 2005 mit dem Thema „How Europeans see Europe. Structure and Dynamics of European Legitimacy Beliefs“ an der Universität von Amsterdam mit Auszeichnung promoviert hat.

Einen besonderen Erfolg haben auch KATHRIN WOHN und TOBIAS HUBERT zu verzeichnen: Beiden ist es gelungen, noch als studentische Mitarbeiter von ZUMA Projekte des Rats für Sozial- und Wirtschaftsdaten zu akquirieren, mit dem für sie schönen „Nebeneffekt“, dass sie für diese Projekte jetzt als wissenschaftliche Mitarbeiter bei ZUMA tätig sind.

KATHRIN WOHN geht in ihrem Projekt über die Effizienz von Weiterbildungsmessung der Frage nach, inwieweit die verschiedenen Dimensionen der Weiterbildung im Mikrozensus, dem Berichtssystem Weiterbildung und der Erhebung zur Betrieblichen Weiterbildung erfasst werden und welche unterschiedlichen Ergebnisse hinsichtlich der Beteiligung verschiedener Bevölkerungsgruppen sich daraus ergeben.

TOBIAS HUBERT betreut im German Microdata Lab als wissenschaftlicher Mitarbeiter das Projekt „Vergleichende Analysen zur beruflichen Weiterbildung mit Mikrozensusdaten“. In diesem Projekt geht es im Schwerpunkt um Determinanten der Weiterbildungsteilnahme. Untersucht werden unter anderem der Einfluss verschiedener sozioökonomischer Faktoren auf die Weiterbildungsaktivität sowie – soweit das im Rahmen von Querschnittsanalysen möglich ist – die Auswirkungen beruflicher Weiterbildung auf Lohneinkommen, Arbeitszeit und berufliche Stellung.

Neu bei ZUMA ist DOROTHÉE BEHR, die nach ihrem Abschluss als Diplom-Übersetzerin an der Universität Heidelberg im European Centre for Comparative Surveys das ESS-Team im Bereich „Übersetzen“ verstärken wird. ZUMA verlassen hat ALEXANDER HAARMANN, bisher wissenschaftlicher Mitarbeiter im ALLBUS.

Und schließlich: Das Projekt „European Social Survey (ESS)“, koordiniert von ROGER JOWELL, City University London, wird auch in der vierten Runde von der EU unterstützt. ZUMA als Mitglied des Central Coordinating Team des ESS ist verantwortlich für Sampling, Questionnaire Design, Translation und Qualitätskontrolle. Um die infrastrukturellen Bedingungen des ESS zu verbessern, wurde das Teilprojekt ESSi (European Social Survey Infrastructure) von der EU mit einem Etat von 6 Millionen Euro für fünf Jahre bewilligt. ZUMA wird dabei hauptsächlich für die Bereiche ‚Improving representativeness and response‘ und ‚Networking activities‘ zuständig sein.

Durchwahl-Rufnummern (Stand: Mai 2006)

Sie erreichen die Mitarbeiter von ZUMA unter der Nummer 0621-1246-(Durchwahlnummer); die Zentrale unter 1246-0. Sie ist von Montag bis Donnerstag von 8.30 bis 17.00 und freitags von 8.30 bis 15.30 besetzt.

Direktion

Prof. Dr. Peter Ph. Mohler (Direktor)	173
Carol Cassidy (Stellv. Direktorin)	146
Margit Bäck (Sekretariat)	172
Elisabeth Bähr (E-Science)	133
Gerrit Kühle (Bibliothek)	159

Interne Infrastruktur

Verwaltung

Dipl.-Kfm. Jost Henze (Leitung)	161
Kerstin Hollerbach, M.A. (KLR)	174

Information & Kommunikation

Kerstin Hollerbach, M.A. (PR & Web-Redaktion)	174
Kai Böge (Intranet & Web-Redaktion)	174

EDV-Infrastruktur

Carol Cassidy (Leitung)	146
-------------------------	-----

Datenbanken

Joachim Wackerow	262
------------------	-----

Wissensvermittlung & Beratung

Wissenschaftlicher Leiter	
Prof. Dr. Ingwer Borg	151
Julia Khorshed (Sekretariat)	274

Projektberater

Dr. Wolfgang Bandilla	136
PD Dr. Michael Braun	176
PD Dr. Jürgen H.P. Hoffmeyer-Zlotnik	175

Dipl.-Soz. Rolf Porst	228
Dr. Beatrice Rammstedt	155
Dagmar Haas (Sekretariat)	152
Patricia Lüder (Workshop-Sekretariat)	221
Christa von Briel (ZUMA-Publikationen)	231

Pretesting

Dipl.-Psych. Peter Prüfer	227
Margrit Rexroth, M.A.	230

Textanalyse, Vercodung

Alfons J. Geis, M.A.	222
Patricia Lüder (Sekretariat)	221

Statistik

PD Dr. Siegfried Gabler	281
Dr. Sabine Häder	282
Dipl.-Math. Michael Wiedenbeck	279

Telefonumfragen

Dipl.-Soz. Michael Schneid	209
Dipl.-Soz. Angelika Stiegler	208

Online-Umfragen

Dipl.-Sozialw. Wolfgang Neubarth	205
----------------------------------	-----

Elektronische Handbücher ZIS/EHES

Dr. Angelika Glöckner-Rist	171
----------------------------	-----

Computerunterstützte Textanalyse

Dipl.-Soz. Juliane Landmann	144
Cornelia Züll	147

Dauerbeobachtung

Wissenschaftlicher Leiter	
PD Dr. Christof Wolf	153
Maria Kreppe-Aygün (Sekretariat)	184

German Microdata Lab

PD Dr. Christof Wolf (Leitung)	153
Dipl.-Soz. Jeanette Bohr	261
Dipl.-Sozialw. Tobias Hubert	186
Dr. Andrea Janßen	266
Dipl.-Sozialw. Yvonne Lechert	271
Andrea Lengerer, M.A.	267
Dr. Paul Lüttinger	268
Dr. Georgios Papastefanou	278
Dipl.-Soz. Bernhard Schimpl-Neimanns	263
Dipl.-Sozialw. Julia Schroedter	264
Dr. Hossein Shahla	253
Dipl.-Soz. Antje Simák	277
Dr. Heike Wirth	269
Irene Fischer (Sekretariat)	265

Soziale Indikatoren

Dr. Heinz-Herbert Noll (Leitung)	241
Dr. Jörg Dittmann	248
Prof. Dr. Caroline Kramer (beurlaubt)	244
Dr. Angelika Scheuer	249
Dr. Stefan Weick	245
Monika Graf (Sekretariat)	242

European Centre for Comparative Surveys

Wissenschaftlicher Leiter	
Prof. Dr. Peter Ph. Mohler	173
Maria Kreppe-Aygün (Assistenz)	184
Julia Khorshed (Sekretariat)	274

ESS-Team

Dipl.-Übers. Dorothee Behr	187
Annelies Blom, M.Phil	272
Dipl.-Verw. Wiss. Matthias Ganninger	189
Dr. Janet Harkness	284
Dipl.-Soz. Achim Koch	280

ALLBUS/ISSP

PD Dr. Michael Braun (Leitung)	176
Dipl.-Soz. Michael Blohm	276
Evi Scholz, M.A.	283
Dipl.-Soz. Martina Wasmer	273

Methodenforschung und -entwicklung

Wissenschaftlicher Leiter	
N.N.	

Drittmittelprojekte**PIONEUR**

Dipl.-Psych. Nina Rother	285
--------------------------	-----

Web Survey Methodology Site (WebSM)

Dipl.-Psych. Lars Kaczmirek	206
-----------------------------	-----

HINWEISE FÜR UNSERE AUTORINNEN UND AUTOREN

Die ZUMA-Nachrichten veröffentlichen – neben Nachrichten des Instituts – Artikel, die von Interesse für die empirische Sozialforschung, insbesondere die Umfrageforschung, sind. Alle Artikel müssen einen methodischen Fokus haben oder zumindest methodische Aspekte stark betonen. Das Spektrum möglicher Artikel ist breit: Es reicht von Grundlagenforschung über angewandte Papiere bis zu Arbeiten, die einen praktisch-operativen Charakter haben. Die Artikel in den ZUMA-Nachrichten sollen für eine breite Leserschaft von Wissenschaftlern und Praktikern im Bereich der empirischen Sozialforschung verständlich sein.

Alle Beiträge, die zur Veröffentlichung in den ZN eingereicht werden, werden von mindestens zwei unabhängigen Gutachtern blind begutachtet.

Die folgenden Regeln sind bei der Abfassung von Manuskripten zu beachten:

1. Manuskripte müssen per email (*porst@zuma-mannheim.de*) eingereicht werden. Der Umfang der Manuskripte (einseitig und 1,5-zeilig beschrieben, Punktgröße 11) soll alles in allem nicht mehr als 30 Seiten betragen.
2. Den Beiträgen sind Abstracts in Deutsch und Englisch (jeweils ca. 15 Zeilen) voranzustellen.
3. Beiträge sind mit dem Dezimalklassifikationssystem zu untergliedern (1 – 2 – 2.1 – 2.2 – 3 usw.). Die Gliederungstiefe geht dabei höchstens auf *eine* Stelle nach dem Punkt.
4. Tabellen enthalten Tabellenummer und Titel im Tabellenkopf, Abbildungen werden analog behandelt.
5. Grafiken sind mittels gängiger Grafiksoftware zu erstellen. Ist eine spezielle Grafiksoftware erforderlich, übernimmt der Autor/die Autorin die endgültige Formatierung der Grafiken in eigener Regie.
6. Bei der Erstellung von Tabellen und Grafiken ist zu berücksichtigen, dass die ZUMA-Nachrichten DIN A5-Format haben und der Satzspiegel 11,5 cm (Breite) x 16 cm (Höhe) beträgt.
7. Anmerkungen und Fußnoten sind mit der Fußnotenfunktion des Schreibprogrammes (im Normalfalle WORD) zu erstellen; bitte nicht gesondert formatieren. Fußnoten sind nur für inhaltliche Kommentare vorzusehen, nicht für bibliographische Hinweise.
8. Literaturhinweise im Text sind nach den folgenden Mustern aufzuführen: Müller (2002) / (Müller 2002) / Müller (2002: 75) / (vgl. Müller 2002: 75) / (Müller 2002; Mayer & Müller 2003).
9. Das Literaturverzeichnis ist gemäß den Richtlinien der American Psychological Association (APA) zu gestalten. Beispiele:
Zeitschriftenbeiträge:
Salzgeber, J. & Stadler, M. (1997). Programm zur Behandlung von Sexualstraftätern. *Zeitschrift für Rechtspsychologie*, 4, 139-141.
Buchveröffentlichungen:
Nicol, A. A. & Pexman, P. M. (1999). *Presenting your findings: A practical guide for creating tables*. Washington, DC: American Psychological Association.
Beiträge in Büchern:
O’Neil, J. M. & Egan, J. (1992). Men’s and women’s gender role journeys: Metaphor for healing, transition, and transformation. In B. R. Wainrib (Hrsg.), *Gender issues across the life cycle* (S. 107-123). New York: Springer.
10. Die Beiträge sind unter Wahrung der gültigen Rechtschreiberegulungen (neue Rechtschreibung) zu erstellen.

AUS DEM INHALT

- Der Rolling Cross-Section Survey – ein Instrument zur Analyse dynamischer Prozesse der Einstellungsentwicklung. Bericht zur ersten deutschen RCS-Studie anlässlich der Bundestagswahl 2005
Rüdiger Schmitt-Beck, Thorsten Faas & Christian Holst 13
- Overreporting: Ein vernachlässigtes Problem und die Schwierigkeit von Gegenmaßnahmen
Sibylle Hardmeier & Marie-Christine Fontana 50
- Zur Wirkung unterschiedlicher materieller Incentives in postalischen Befragungen. Ein Literaturbericht
Fred Berger 81

ISSN 0941-1670