

methoden daten analysen

ZEITSCHRIFT FÜR EMPIRISCHE SOZIALFORSCHUNG

mda

2007, Jahrgang 1, Heft 2



Wolfgang Sodeur Entscheidungsspielräume von Interviewern bei der Wahrscheinlichkeitsauswahl

Rolf Becker, Regula Imhof und Guido Mehlkop Monetäre Anreize bei einer postalischen Befragung und Antworten auf Fragen zur Delinquenz

Stefan Hunsicker und Yvonne Schroth Die Kombination von Mobilfunk- und Festnetzstichproben

Henning Best Die Messung von Nutzen und subjektiven Wahrscheinlichkeiten

Herausgegeben von *Christof Wolf
Marek Fuchs
Bärbel Knäuper
Steffen Kühnel*

Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung

Die Zeitschrift wird herausgegeben von der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS). Die GESIS ist eine Einrichtung der Leibniz-Gemeinschaft.

Herausgeber: Christof **Wolf** (Mannheim, geschäftsführend), Marek **Fuchs** (Kassel), Bärbel **Knäuper** (Montreal), Steffen **Kühnel** (Göttingen)

Wissenschaftlicher

Beirat: Hans-Jürgen **Andreß** (Köln), Andreas **Diekmann** (Zürich), Sabine **Häder** (Mannheim), Udo **Kelle** (Marburg), Dagmar **Krebs** (Giessen), Frauke **Kreuter** (College Park, Maryland), Edith **de Leeuw** (Utrecht), Norbert **Schwarz** (Ann Arbor)

Redaktion: Paul Lüttinger
GESIS-ZUMA
Postfach 12 21 55
68072 Mannheim
Tel.: 0621 – 1246-268
E-Mail: mda@gesis.org
Internet: www.gesis.org/Publikationen/Zeitschriften/MDA/

Die MDA deckt alle Fragestellungen aus dem Bereich der Empirischen Sozialforschung ab, insbesondere aus dem Bereich der Umfragemethodik. Im Vordergrund stehen Artikel, welche die methodischen und/oder statistischen Kenntnisse der Profession erweitern, sowie Beiträge, die sich mit der Anwendung der Methoden der Empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder solche, in denen ein statistisches Verfahren exemplarisch angewandt wird. Obwohl der Schwerpunkt auf Umfragemethoden liegt, sind Beiträge zu anderen methodischen Bereichen willkommen.

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

Der Nachdruck von Beiträgen ist nach Absprache möglich. Die MDA erscheint dreimal im Jahr und steht als Printversion und online zur Verfügung. Die Registrierung für den Bezug der MDA erfolgt über die Web-Seiten der GESIS:

http://www.gesis.org/Publikationen/Zeitschriften/MDA/MDA_subscribe.asp

Druck: Concordia-Druckerei König oHG, Mannheim-Sandhofen
Gedruckt auf chlorfrei gebleichtem Papier.

ISSN 1864-6956

1. Jahrgang 2007 © GESIS, Mannheim, Dezember 2007

Inhalt

105 Editorial

FORSCHUNGSBERICHTE

- 107 Entscheidungsspielräume von Interviewern bei der Wahrscheinlichkeitsauswahl. Ein Vergleich von ALLBUS-Erhebungen
Wolfgang Sodeur
- 131 Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz. Empirische Befunde eines Methodenexperiments
Rolf Becker, Regula Imhof und Guido Mehlkop
- 161 Die Kombination von Mobilfunk- und Festnetzstichproben. Eine praktische Anwendung des Dual-Frame-Ansatzes
Stefan Hunsicker und Yvonne Schroth
- 183 Die Messung von Nutzen und subjektiven Wahrscheinlichkeiten. Ein Vorschlag zur Operationalisierung der Rational Choice Theorie
Henning Best
-

REZENSIONEN

- 213 Offene Interviews führen und auswerten
Sammelbesprechung
Nina Baur
- 220 Funktionale Äquivalenz in interkulturell vergleichenden Umfragen. Mythos und Realität. Michael Braun, 2006
Franz Höllinger
- 223 The Telephone Interviewer's Handbook. How To Conduct Standardized Conversations. Patricia A. Gwartney, 2007
Frank Faulbaum
- 224 Empirische Sozialforschung. Eine Einführung. Michael Häder, 2006
Jost Reinecke

ANKÜNDIGUNGEN

- 227 G.O.R. 08 - 10th International G.O.R. Conference. General Online Research
- 228 Workshop: Beginning to Advanced Methods in Structural Equation Modeling
- 229 Workshop: Einführung in familien- und haushaltsbezogene Analysen mit dem Mikrozensus
- 230 Call for Papers: Interims Meeting of the ISA Research Committee 33 (RC33) on „Logic and Methodology in Sociology“
-
- 232 AutorInnen, RezensentInnen, GutachterInnen 2007
- 233 Hinweise für unsere Autorinnen und Autoren

Editorial

Liebe Leserin, lieber Leser,

mit diesem zweiten Heft ist der erste Jahrgang der MDA abgeschlossen. Insgesamt wurden in diesem Jahr 27 Artikel zur Veröffentlichung eingereicht. Dies entspricht in etwa der Zahl der Beiträge, die in den letzten Jahren ihres Erscheinens den ZUMA-Nachrichten und den ZA-Informationen gemeinsam jährlich zur Veröffentlichung angeboten wurden. Von den 27 eingereichten Artikeln wurden acht Beiträge nach intensiver Begutachtung publiziert, 16 Beiträge wurden abgelehnt, und über drei Beiträge wurde noch nicht abschließend entschieden. Unser Dank gilt allen KollegInnen, die ihre Manuskripte zur Publikation bei der MDA eingereicht haben sowie den vielen GutachterInnen, die mit ihren sehr sorgfältig erstellten Expertisen spürbar zur Qualität der Zeitschrift beitragen. Eine vollständige Liste aller AutorInnen, RezensentInnen und GutachterInnen finden Sie am Ende dieses Heftes.

Die Beiträge des ersten Jahrgangs beziehen sich auf klassische Fragen der Umfrageforschung, die sich jedoch durch Veränderungen gesellschaftlicher Umstände immer wieder neu stellen. So etwa die Erhöhung des Rücklaufs bei postalischen Befragungen, Probleme der Stichprobenziehung für spezielle Populationen (hier konkret Aussiedler) bzw. bei sich verändernden Rahmenbedingungen (hier die Integration von Festnetz- und Mobilfunkstichproben) oder die Auswirkungen unterschiedlich strikter Intervieweranweisungen auf die Repräsentativität von Stichproben.

Neben den Originalbeiträgen spielen die Rezensionen in der MDA eine wichtige Rolle, da sie einen Überblick über einschlägige Neuerscheinungen geben. Dabei sind wir für Hinweise auf Publikationen, die sich für eine Besprechung in der MDA eignen, jederzeit dankbar.

Zum Abschluss möchten wir Sie noch über eine personelle Veränderung informieren. Rolf Porst wird seine Tätigkeit als Redakteur der MDA zum Ende des Jahres niederlegen. Seinem Einsatz ist es zu verdanken, dass die Zeitschrift ihre Arbeit ohne größere Schwierigkeiten aufgenommen hat. Für seine Mitwirkung

danken wir ihm herzlich; wir werden seine Erfahrung, sein Augenmaß und seine Anregungen vermissen. Wir freuen uns jedoch, dass wir mit Paul Lüttinger wiederum einen ebenso erfahrenen und lang gedienten Redakteur als Nachfolger gewinnen konnten. Für unsere Autoren wird sich damit nichts ändern. Sie erreichen uns nach wie vor unter der Adresse mda@gesis.org.

Wir wünschen Ihnen ein erfolgreiches neues Jahr.

CHRISTOF WOLF * MAREK FUCHS * BÄRBEL KNÄUPER * STEFFEN KÜHNEL

Entscheidungsspielräume von Interviewern bei der Wahrscheinlichkeitsauswahl

Ein Vergleich von ALLBUS-Erhebungen

Interviewers' scope for influence on random sampling

A comparative analysis of ALLBUS surveys

Wolfgang Sodeur

Zusammenfassung

Eine ideale Wahrscheinlichkeitsauswahl würde genau festgelegten Regeln folgen. Beteiligte Personen müssten sich exakt an diese Regeln halten. Reale Wahrscheinlichkeitsauswahlen in der Sozialforschung jedoch werden auf der letzten Stufe meist von Interviewern durchgeführt, denen je nach Verfahren unterschiedliche Verhaltensspielräume entweder explizit zugestanden oder mangels hinreichender Kontrollen zumindest nicht verwehrt werden. Abhängig vom faktisch vorhandenen Spielraum benachteiligen Interviewer die schwerer erreichbaren Personen aufgrund des relativ höheren Aufwandes zur Kontaktaufnahme. Die dabei erzeugte "Wahrscheinlichkeitsauswahl" ist verzerrt: Schwerer erreichbare Personen fehlen systematisch in umso größerem Umfang, je größer der Spielraum der Interviewer bei der Auswahl war.

Diese Annahmen wurden anhand der 12 ALLBUS-Erhebungen von 1980-2000 überprüft: Bei direkter Auswahl von Personen aus Einwohnermeldeamtslisten (geringer Spielraum) fanden wir einen viel höheren Anteil der relativ schwer erreichbaren Personen als in

Abstract

Looking for hidden bias in random samples, we re-analyzed 12 ALLBUS Surveys (Germany's GSS, 1980-2000). In these surveys people had been selected by varying procedures: the interviewers had been allowed some amount of scope for own activities and thereby (hypothesis) induced bias of differing severeness.

The sampling bias is measured by means of internal criteria. By definition, the total population as well as the random sample are restricted to the end that known population parameters arise, in this case a total population of couples (one men, one women, same household, at least one child up to 14 years old), where the male person is working full-time outside the household and the female is not. As a result of these restrictions, an unbiased random sample is expected to consist of 50% women exactly, in spite of them being more easily accessed than their male partners.

Comparing the results of existing random samples (ALLBUS) to the expected values, the relative frequency of women differs systematically: in cases of sampling procedures

haushaltsbezogenen Auswahlen wie z. B. nach dem ADM-Design (relativ großer Spielraum).

Die Spielräume werden als partielle Rangordnung aus den Methodenberichten der 12 ALLBUS-Erhebungen abgeleitet. Die Auswahlverzerrungen zeigen sich durch den Grad der Abweichung von bekannten Verteilungsparametern, die definitorisch allein durch die Abgrenzung einer geeigneten Teilgesamtheit erzeugt werden.

prescribing the interviewers behavior in detail, we found a significant higher number of males (difficult to access) compared to samples which allowed more freedom.

1 Einführung: Befragungsbedingungen, Interviewerverhalten und Auswahlresultate¹

Wahrscheinlichkeitsauswahlen folgen zunächst theoretischen Konzepten. Versuche zur praktischen Umsetzung dieser Konzepte zielen auf eine möglichst große Annäherungen an das Ideal. Jedoch verfolgen sie dieses Ziel nicht bedingungslos, sondern müssen auch die entstehenden Kosten und den Bedarf an Zeit im Zusammenhang mit den verfügbaren Mitteln beachten.

Wie bei allen Kompromissen geraten Forscher auch hier in ein Spannungsfeld zwischen den erwünschten Eigenschaften der zu erhebenden Daten wie z. B. der Kalkulierbarkeit des Risikos bei ihrer Übertragung auf angebbare Grundgesamtheiten auf der einen und den verfügbaren Ressourcen an Geld und Erhebungszeit auf der anderen Seite.

Dilemmata der hier angedeuteten Art sind allgemein bekannt. Völlig aufzulösen wären sie nur unter den unrealistischen Bedingungen unbegrenzter Zeit- und Geldmittel sowie beliebiger Möglichkeiten zur Einflussnahme auf die Auswahl- und Erhebungsprozesse. Große Erhebungsinstitute bieten deshalb gezielt Mischlösungen² an, die auf die tatsächlichen oder vermeintlichen Bedürfnisse vermuteter Auftraggeber zugeschnitten sind. Damit werden einige Elemente einer

1 Der Aufsatz wurde ursprünglich eingereicht und angenommen für eine geplante Festschrift, die 2003 zum 70. Geburtstag von Erwin K. Scheuch erscheinen sollte. Diese Festschrift wird posthum nicht mehr erscheinen. Der Aufsatz bleibt dem Andenken an Erwin K. Scheuch gewidmet. Für Anmerkungen zu einer früheren Fassung danke ich Karl-Heinz Reuband und einem anonymen Gutachter.

2 Z. B. Auswahlen nach dem Random Route-Verfahren, bei denen erleichterte Auswahlbedingungen für die Interviewer zu Zeit- und Kosteneinsparungen führen, mögliche Auswahlverzerrungen dadurch aber mithilfe einer zusätzlichen Quotensteuerung begrenzt werden sollen.

Wahrscheinlichkeitsauswahl realisiert und gleichzeitig andere bewusst aufgegeben. Bei unbestrittener Notwendigkeit solcher Kompromisse erscheint die mangelnde Transparenz ihrer Folgen als problematisch.

Als potentiell besonders zeit- und kostenaufwändig erweist sich bei einer Wahrscheinlichkeitsauswahl der Umgang mit der unterschiedlichen Erreichbarkeit der Auswahleinheiten.³ Zur Illustration seien dazu zwei extreme Beispiele mit besonders hohem bzw. niedrigem Zeit- und damit Kostenaufwand beschrieben. Gemeinsam sei den Beispielen ein mehrstufiges Auswahlverfahren, das auf der 1. Stufe aus einer Gebietsauswahl nach dem ADM-Stichprobendesign (ADM/AG.MA 1999) besteht. Auf den folgenden Auswahlstufen nimmt ein Interviewer innerhalb des ihm zugewiesenen Gebiets von einem zufällig festgelegten Anfangspunkt aus einen Zufallsweg („random route“), dem entlang zunächst Häuser und darin jeweils Wohnungen bzw. Privathaushalte nach vorher festgelegten Regeln in eine bestimmte Reihenfolge gebracht werden. Aus dieser geordneten Reihe von Haushalten werden dann aufgrund einer systematischen Wahrscheinlichkeitsauswahl zunächst Haushalte und schließlich aus diesen Haushalten jeweils eine durch Wahrscheinlichkeitsauswahl (mittels Schwedenschlüssel) ermittelte Zielperson zur Befragung ausgewählt.

Probleme bei der Realisierung können auftreten,

- (a) weil entweder im ausgewählten Haushalt keine Person angetroffen wird oder
- (b) die Mitwirkung bei der Aufstellung einer Liste aller zum Haushalt gehörenden Personen aus der Grundgesamtheit verweigert wird,
- (c) weil die innerhalb des Haushalts aus der zuvor erstellten Liste ausgewählte Person nicht anwesend ist oder
- (d) die Zielperson die Befragung verweigert.

Die Realisationsversuche sollten nach allgemein anerkannten Regeln mit der ersten Verweigerung (b) oder (d) enden. Gestaltungsoptionen bleiben damit nur noch für die Fälle der Nicht-Erreichbarkeit (a) und (c). Von der Art der Behandlung dieser Fälle hängt der zeitliche und finanzielle Aufwand bei der Erhebung ebenso ab wie der Grad der Annäherung der realisierten Stichprobe an eine Wahrscheinlichkeitsauswahl.

(1) Variante mit hohen Kosten und relativ großer Näherung an eine Wahrscheinlichkeitsauswahl: Zur Durchführung von maximal 10 Befragungen wählt der Interviewer nach den oben genannten Regeln 10 Haushalte und darin je eine Person aus und versucht, notfalls unter Wiederholung von Kontaktversuchen, möglichst

3 Zum Problem der Erreichbarkeit vgl. Schnell (1997, Kap. 5).

viele der Befragungen zu realisieren. Ersatzadressen für nicht realisierbare Befragungen stehen ihm nur unter restriktiven Bedingungen (für stichprobenneutrale Ausfälle) zur Verfügung.

(2) Variante mit relativ geringen Kosten und starker Abweichung von einer Wahrscheinlichkeitsauswahl: Zur Durchführung von maximal 10 Befragungen darf der Interviewer 30 Haushalte bzw. Personen (siehe (1)) auswählen. Die Reihenfolge der Kontaktversuche steht ihm ebenso frei wie Wiederholungen dieser Versuche.

Das letztgenannte Verfahren erlaubt eine relativ *kosteneffiziente* Durchführung der Befragungen, weil zeitraubende Wiederholungen von Kontaktversuchen und sogar die Wahrnehmung in Aussicht gestellter Befragungen mit ausgewählten, aber augenblicklich abwesenden Haushalts-Mitgliedern nur realisiert werden müssen, wenn die angestrebten 10 Interviews nicht auf anderem, zeitsparenderem Wege erzielt werden können. Es gibt Gründe zu befürchten, dass in diesem Fall die tatsächliche Auswahl im Hinblick auf eine unterschiedliche Erreichbarkeit der Zielpersonen gegenüber einer reinen Wahrscheinlichkeitsauswahl systematisch verzerrt ist.⁴

Das Ziel dieser Arbeit ist eine exemplarische Untersuchung der Zusammenhänge zwischen den jeweils festgelegten Auswahlregeln und dem Umfang der Auswahlverzerrungen. Damit wird ein wichtiger Grund für die systematischen Verzerrungen bei der Wahrscheinlichkeitsauswahl im Prozess der Auswahlhandlungen gesucht (Scheuch 1974: 69). Als empirische Basis dienen die 12 ALLBUS-Erhebungen aus den Jahren 1980-2000.

In Abschnitt 2 werden zunächst die Methodenberichte dieser 12 Erhebungen mit dem Ziel analysiert, den Umfang des Spielraums der Interviewer bei der Auswahl der zu befragenden Personen und bei der Realisierung der Befragungen mit den ausgewählten Personen einzustufen.

Abschnitt 3 beschreibt ein Verfahren zur Kennzeichnung von Auswahlverzerrungen anhand *interner Kriterien*, das keine empirischen Vorkenntnisse über die Grundgesamtheit benötigt. Damit wird die Anzeige spezieller Verzerrungen möglich, die das Interviewerverhalten im eigenen Interesse bei der Auswahl und Befragung unterschiedlich leicht erreichbarer Personen erzeugt.

Entsprechende Hypothesen über den Zusammenhang zwischen unterschiedlich großen Spielräumen der Interviewer und dem Umfang dadurch erzeugter Auswahlverzerrungen werden im abschließenden Abschnitt 4 mit den Daten der 12 ALLBUS-Erhebungen überprüft.

4 Vgl. dazu die Diskussion um die Berechnung von Korrektur-Gewichten zur Antreffbarkeit in den ALLBUS-Methodenberichten u. a. von 1991 (Abschnitt 6.3.1 Gewichtungen) und 1992 (Abschnitt 4.2.3).

2 Varianten praktizierter Auswahlverfahren

In diesem Abschnitt beschreiben wir kurz diejenigen Aspekte der Stichprobenziehung verschiedener ALLBUS-Erhebungen, die für die hier angestrebte Analyse bedeutsam sind. Wie bereits einleitend dargestellt, handelt es sich dabei um die Regeln bei der Auswahl auf der zweiten und gegebenenfalls dritten Stufe, nämlich bei der entweder direkten Auswahl von Personen oder der Auswahl von zunächst Haushalten und anschließend Personen innerhalb der Haushalte.

Leider sind die Beschreibungen dieser Stufen des Auswahlprozesses in den Methodenberichten zu den ALLBUS-Erhebungen nicht ähnlich detailliert wie die Beschreibungen der Gebietsauswahlen auf der ersten Stufe. Vor allem fehlen solche Einzelheiten über die Rahmenbedingungen der Erhebung, die Rückschlüsse auf das mögliche Verhalten der Interviewer innerhalb ihrer Gestaltungsspielräume – und damit indirekt auch die Einschätzung der Gefahr von Regelverstößen – erlauben würden: Zu nennen ist hier einmal die Höhe der Zahlungen für ein realisiertes Interview sowie gegebenenfalls die Höhe zusätzlicher Zahlungen für wiederholte Kontaktversuche, ferner die Art, wie in den Instituten die Einhaltung der vorgeschriebenen Verfahrensweisen kontrolliert wird. Dieses betrifft sowohl die laufende Erhebung wie auch frühere Erhebungen des Instituts, die auf dem Umweg über die Kenntnisse der Interviewer als *internalisierte Kontrollen* wirken könnten.

2.1 Bezahlungsform der Interviewer und Folgen für die Auswahl

In aller Regel werden Interviewer in Deutschland durch die Institute auf der Basis von Stückkosten entlohnt, d. h. durch einen festen Betrag für jede realisierte Befragung. Diese Form der Entlohnung führt zu leicht kalkulierbaren Erhebungskosten und erfordert weder weitergehende Kontrollen des zeiteffizienten Verhaltens der Interviewer noch besondere Maßnahmen zur Kostenkontrolle während der Erhebung.

Andererseits tragen bei dieser Form der Entlohnung, d. h. ausschließlich über den Stücklohn pro Befragung, die Interviewer das alleinige Risiko für jeden zusätzlichen Aufwand, der zum Beispiel durch eine wiederholte Nicht-Erreichbarkeit von Haushalten oder der im Haushalt ausgewählten Befragungspersonen zustande kommt oder auch durch besonders lange Befragungszeiten, die in der Person des Befragten oder der besonderen Datenlage begründet sind. Es ist davon auszugehen, dass Interviewer diese finanziellen Bedingungen beachten und ihr Verhalten auch an der erwarteten Höhe der Entlohnung orientieren.

Wenn die Bezahlung der Interviewer ausschließlich auf dem Stücklohn beruht, ist ihre Entlohnung (*ceteris paribus*) am relativ höchsten, wenn die Befragung unmittelbar im Rahmen des Zufallsweges durch das Gebiet geschieht. Das gilt für die Aufnahme der Haushalts-Liste, die Auswahl des zu befragenden Haushalts, den ersten Kontakt mit dem Haushalt, die Aufnahme der Liste aller zur Grundgesamtheit gehörenden Haushalts-Mitglieder und die Auswahl der zu befragenden Person (z. B. mittels Schwedenschlüssel). Bezogen auf jede einzelne Erhebungseinheit kann diese – mit optimaler Entlohnung verbundene – Bedingungskonstellation von Interviewern nicht beeinflusst werden. Anders verhält es sich aber mit einer Reihe entsprechender Erhebungseinheiten auf dem Zufallsweg durch das Gebiet: Da die auszuwählenden Haushalte in der Regel im Erhebungsgebiet dicht beieinander liegen, kosten sequentielle Kontaktversuche wenig Zeit und alle in derselben Sequenz realisierten Befragungen können als optimal entlohnt gelten.

Bereits schlechter sind die Realisationschancen für Befragungen in einem unmittelbar auf die erste Sequenz folgenden zweiten oder dritten Durchgang durch das Gebiet, wenn aufgrund einiger bereits realisierter Befragungen weniger Erhebungseinheiten zur Auswahl stehen. Sehr viel ungünstiger wird die Bezahlung jedoch, wenn Kontakte mit Haushalten oder Befragungen bereits ausgewählter Zielpersonen an anderen Tagen nach wiederholtem Besuch des Erhebungsgebiets versucht werden. Die erwartete Entlohnung wird im Extremfall u. U. negativ, wenn an einem Ort nur einzelne Haushalte übrig bleiben oder aufgrund vorangehender Erfahrungen nur wenig Hoffnung besteht, in den ausgewählten Haushalten jemanden anzutreffen.

Interviewer, welche die erwartete Höhe ihrer Entlohnung verhaltenswirksam beachten, werden sich also mit unterschiedlicher Intensität um die Realisierung von Befragungen bemühen. Sie werden ihr Bemühen ganz einstellen,

- (a) wenn die erwartete Entlohnung eine Mindestgröße unterschreitet oder
- (b) wenn Alternativen zur Durchführung anderer Befragungen unter günstigeren Bedingungen innerhalb derselben Erhebung oder
- (c) in einer anderen Erhebung bestehen.

Zur erst- und letztgenannten Bedingung (a,c) liegen uns keine Informationen über die untersuchten Erhebungen vor. Wir müssen daher annehmen, dass sie zwischen den zu vergleichenden ALLBUS-Erhebungen nicht variieren. Es gibt aber Zweifel an dieser Annahme: Die Interviewerstäbe der mit der Erhebung beauftragten Institute unterscheiden sich deutlich hinsichtlich ihrer Zusammensetzung nach Geschlecht und Alter und sind nicht einmal innerhalb eines Instituts identisch über die Zeit (vgl. Methodenberichte zu den ALLBUS-Erhebungen).

Bedingung (b) dagegen kann anhand der vorliegenden Beschreibungen beurteilt und zum Vergleich zwischen den Erhebungen genutzt werden. Im folgenden Abschnitt wird deshalb zusammengestellt, in welchem Umfang innerhalb der festgelegten Regeln Spielräume der Interviewer bei der Auswahl der Haushalte und damit Gestaltungsmöglichkeiten bei der Festlegung potentiell zu kontaktierender Haushalte bestehen.

Grundsätzlich kann jeder Interviewer seine Tätigkeit mehr oder weniger intensiv betreiben und – bei Erwartung unzureichender Entlohnung – jederzeit ganz einstellen. In diesem Sinne müssen auch Ausschöpfungsraten der Bruttostichprobe nicht notwendig den Befragten („Verweigerungen“) angelastet werden, sondern könnten teilweise auch Folgen der Abbruchentscheidungen der Interviewer aufgrund unzureichender Bezahlung von Befragungen bei erwarteten Mehrfachkontakten sein.

Natürlich gibt es prinzipiell auch die Möglichkeit der Regelverstöße, in diesem Fall also der Kontaktierung eigentlich nicht ausgewählter Haushalte, die tatsächlich oder vermeintlich leichter erreichbar sind und/oder der Befragung eigentlich nicht als Zielpersonen in einem Haushalt ausgewählter Personen. Solche Verstöße sind nur unter zusätzlichen Kontrollen aufzudecken, deren Umfang jedoch innerhalb der jeweiligen ALLBUS-Erhebung nicht immer klar dokumentiert ist.

Zusätzliche Kontrollen wären aber auch erforderlich, wenn Interviewer eine ergänzende, der Befragungszeit entsprechende Entlohnung ihres Zeitaufwandes für mehrfache Kontaktversuche eines Haushalts oder für die wiederholten Besuche eines Haushalts zur Befragung der ausgewählten Zielperson erhielten. Andernfalls wären die Erhebungskosten nicht im Vorhinein kalkulierbar und (unredliche) Interviewer erhielten die Möglichkeit zur direkten Manipulation ihres Einkommens. Alternativen zum reinen Stücklohn für realisierte Interviews durch zusätzliche, den Zeitaufwand berücksichtigende Elemente werden deshalb in der Erhebungspraxis nur äußerst selten gewählt und kommen unter den hier analysierten ALLBUS-Erhebungen unseres Wissens nicht vor.

Die genannten Überlegungen zum rationalen Erwerbsverhalten der Interviewer fassen wir in einer Hypothese zusammen:

Die Auswahl von Haushalten und darin von Zielpersonen durch die Interviewer wird umso größere Selektionseffekte aufgrund unterschiedlicher Erreichbarkeit aufweisen, je größer der Spielraum der Interviewer bei der Auswahl ist.

Ungeprüfte bzw. vereinfachende, der Realität nicht immer entsprechende Annahmen sind dabei insbesondere:

- Die Interviewerstäbe in allen ALLBUS-Erhebungen sind gleich hinsichtlich ihrer zeitlichen Disposition (z. B. Halb-/Ganztagstätigkeit als Interviewer, Verteilung der Kontakt- und Befragungsversuche auf Tageszeiten).
- Die Interviewerstäbe in allen ALLBUS-Erhebungen sind gleich hinsichtlich ihrer Berufserfahrungen als Interviewer, ihrer professionellen Ethik usw.
- Die Entlohnungsformen (Verhältnis von Stücklohn für Befragung versus zusätzlicher Entlohnung für besonderen zeitlichen Aufwand) sind bei allen Erhebungen gleich.
- Bei allen Erhebungen wurden in gleichem Umfang die Einhaltung der Auswahlregeln kontrolliert bzw. es bestanden gleiche Erwartungen der jeweiligen Interviewerstäbe gegenüber entsprechendem Kontroll-Verhalten.

Sowohl die hier verfolgte Hypothese zum Verhalten der Interviewer wie leider auch Zweifel an der Richtigkeit der vereinfachenden Hilfsannahmen werden durch einige frühere Analysen gestützt.

Diese Analysen leiden allerdings genau wie unsere eigene Studie unter einem Mangel an Daten. Die eigentlich benötigten Informationen über die Form der Bezahlung – dies trifft insbesondere den Fall von wiederholten Versuchen zur Kontaktaufnahme – und über den genauen Ablauf der Interviewer-Einweisung und -Kontrolle liegen in keinem Fall vor. In nur wenigen Fällen gibt es zumindest Daten mit indirektem Bezug zu den hier diskutierten Fragen, die zudem – mit Ausnahme vielleicht der Studie von Alt/Bien/Krebs (1991) – nur entweder Eigenschaften der Erhebung eines Instituts oder nur Eigenschaften der Interviewer bzw. ihres Verhaltens betreffen. Eine Interpretation im Sinne der hier diskutierten Fragen ist dann wieder nur unter ganz ähnlichen wie den oben genannten oder unter dazu komplexeren Hilfsannahmen möglich.

Schnell (1997: 58-59) verweist in seiner umfangreichen Untersuchung des „Nonresponse in Bevölkerungsumfragen“ auf den Einfluss des Ermessensspielraums, der Interviewern bei der Auswahl der zu Befragenden eingeräumt wird. Bis zu diesem Zeitpunkt lagen nur wenige Studien mit Einwohnermeldeamts-Stichproben (EWM) vor, bei denen ein „kleinerer Einfluss des Interviewers auf die Auswahl der Zielperson und die Art der Ausfälle erwartet“ wird (Schnell 1997: 59). Sie zeigten eine deutliche Verminderung des Anteils der nicht-erreichbaren Personen gegenüber vergleichbaren Auswahlen nach dem meist gewählten ADM Design (Koch/Gabler/Braun 1994; Alt/Bien/Krebs 1991).

In seiner eigenen Analyse von Erhebungsberichten aus 300 Bevölkerungsumfragen, die im Zentralarchiv für empirische Sozialforschung archiviert sind, findet er starke Unterschiede in der gesamten Ausschöpfungsquote wie auch im Anteil der

Verweigerungen und der Ausfälle aufgrund Nicht-Erreichbarkeit sowohl *zwischen den Instituten* wie auch über die Zeit (Schnell 1997: 76-100). Leider werden damit die oben genannten Zweifel an der Tragfähigkeit unserer Hilfsannahmen weiter gestärkt.

Über die starke Variation der Ausschöpfungsergebnisse zwischen Erhebungsinstituten hinaus berichtet Schnell auch über – zumindest indirekte – Hinweise auf die Folgen unterschiedlicher Ermessensspielräume der Interviewer. Insbesondere die Bevölkerungsumfragen im Rahmen der Media-Analysen zeigen über Institutsgrenzen hinweg eine höhere Ausschöpfung bzw. einen geringeren Anteil an Verweigerern und Nicht-Erreichbaren als andere Umfragen mit kommerziellen oder wissenschaftlichen Auftraggebern. Mangels direkter Informationen über die Kontrollen im Prozessverlauf der Erhebungen können diese Ergebnisse zwar nicht eindeutig den hier behandelten Spielräumen der Interviewer zugeschrieben werden. Es gibt aber plausible Erklärungen, warum vor allem die Bevölkerungsumfragen der Media-Analyse mit einem relativ hohen Kontrollpotential im Erhebungsprozess verbunden sein könnten:

„Die Media-Analyse ist hoch standardisiert und von der wahrgenommenen Güte der Feldarbeit hängt die weitere Zukunft des Umfrageinstituts innerhalb der Media-Analyse ab.“ Deshalb „sollten die wiederholten Ergebnisse eines Instituts nicht zu stark von den Ergebnissen der anderen Institute abweichen.“ (Schnell 1997: 79).

Auch innerhalb der Erhebungen zur Media-Analyse stellt Schnell jedoch (trotz der besonderen Standardisierung) große Unterschiede bei der Verteilung der Zahl der Kontaktversuche bis zur Realisierung des Interviews zwischen den Instituten fest:

„Damit muss für die Media-Analyse festgehalten werden, dass der Einfluss der Erhebungsinstitute auf die berichtete Erreichbarkeit der Befragten größer ist als der Einfluss der Befragten. Entweder unterscheidet sich also das Kontaktverhalten der Interviewer oder die Art der Berichterstattung zwischen den Instituten.“ (Schnell 1997: 223).

An anderer Stelle geht Schnell anhand einzelner Erhebungen (ALLBUS 1980,1988; vgl. Schnell 1997: 229-236)⁵ mit entsprechend günstiger Datenlage über das Interviewer-Verhalten auch einigen der möglichen Ursachen für unterschiedliche Anteile der Nicht-Erreichbaren nach. So vergleicht er z. B. die insgesamt für den ALLBUS 1988 benötigte Feldzeit von 10 Wochen mit entsprechenden

5 Die in unserer Arbeit analysierten ALLBUS-Erhebungen enthalten ebenfalls diese beiden Datensätze von 1980 und 1988. Umgekehrt sind die von uns analysierten ALLBUS-Erhebungen fast sämtlich (d. h. bis 1994) auch in den von Schnell analysierten 300 archivierten Studien enthalten.

Feldzeiten auf der Mikroebene, d. h. der von einzelnen Interviewern an bestimmten Sample-Points für ihre Befragungen verbrachten Zeit (Schnell 1997: 229-231). In mehr als der Hälfte aller Fälle beschränkt sich diese individuelle Feldzeit an einem Sample-Point auf etwa 3 Tage.

Wiederholte Kontaktversuche werden darüber hinaus (nun auf Basis der Kontaktprotokolle des ALLBUS 1980 ermittelt) zum größten Teil in kurzem zeitlichen Abstand zum ersten Kontaktversuch unternommen: „2/3 der Fälle weisen eine Differenz von maximal 4 Stunden auf, nur ca. 7% der Fälle besitzen eine Differenz von mehr als 8 Stunden“ (Schnell 1997: 233).

Beide Ergebnisse sind im Sinne der hier verfolgten Hypothese als suboptimale Verhaltensweisen der Interviewer zu interpretieren, die möglicherweise durch strengere Vorgaben und Kontrollen der Erhebungsinstitute mit dem Ziel einer höheren Ausschöpfung verändert werden könnten.

Wir müssen aber nochmals darauf hinweisen, dass die Datenlage bei diesen letztgenannten Untersuchungen in positiver wie negativer Hinsicht komplementär ist zur Datenlage sowohl bei unserer eigenen wie auch der vorher genannten Untersuchung der 300 Erhebungsberichte durch Rainer Schnell (1997). Während dort die Variation der Ausschöpfung etc. nur auf der Makro-Ebene, d. h. über Erhebungen bzw. Institute geprüft wird, betreffen die zuletzt berichteten internen Verhältnisse in den ALLBUS-Erhebungen von 1980 und 1988 nur die Mikro-Ebene des Interviewerverhaltens unter den jeweils relativ konstanten Organisations- und Kontrollbedingungen allein einer Erhebung eines Instituts.

Etwas günstiger ist die Datenlage bei der abschließend diskutierten Untersuchung des Deutschen Jugendinstituts zum Thema ‚Partnerschaft und Familie heute‘ (Alt/Bien/Krebs 1991). Zur Prüfung der Konsequenzen unterschiedlicher Verfahren zur Wahrscheinlichkeitsauswahl wurde die Stichprobe teilweise nach dem ‚random route-Verfahren‘ (N=7.032) und teilweise aus den EWM-Registern (N=3.011) gezogen. Im ersten Schritt des mehrstufigen Auswahlverfahrens folgten beide Teilstichproben dem ADM-Design. Für die hier behandelten Fragen ist die Studie vor allem deshalb interessant, weil dasselbe Erhebungsprogramm zeitgleich in vergleichbaren Regionen vom gleichen Erhebungsinstitut mit zwei unterschiedliche Auswahlverfahren durchgeführt wurde.

Vergleiche der Verteilungen nach Geschlecht und Erwerbstätigkeit zeigen deutliche Unterschiede zwischen beiden Teilstichproben, wobei die EWM-Stichprobe besser mit den Ergebnissen des Mikrozensus übereinstimmt als die random route Stichprobe. Letztere enthält viel weniger Männer und Erwerbstätige. Als Gründe dafür vermuten Alt, Bien und Krebs, dass den Interviewern im Random Route-Ver-

fahren zumindest faktisch ein größerer Ermessensspielraum zur Verfügung stand. Folgerichtig suchen sie nach „Indikatoren für den unterschiedlichen Zugang zum Feld in den beiden Stichprobenverfahren“ (Alt/Bien/Krebs 1991: 67).

Während in der Random Route-Stichprobe die meisten Befragungen beim ersten Kontaktversuch realisiert werden, wird dieser Gipfel bei der EWM-Stichprobe erst im zweiten Versuch erreicht. Nach beiden Verfahren werden im ersten Versuch (gegenüber dem Mikrozensus) mehr Frauen und Nicht-Erwerbstätige befragt, also „leicht Erreichbare“ (S. 69). In beiden Stichproben wird diese Verzerrung mit zusätzlichen Kontaktversuchen geringer. Diese Tendenz ist aber in der EWM-Stichprobe stärker. Außerdem werden in der EWM-Stichprobe insgesamt viel mehr Befragte nach zwei oder mehr Kontaktversuchen befragt als in der Random Route-Stichprobe.

Alt, Bien und Krebs interpretieren den größeren Ermessensspielraum der Interviewer beim Random Route-Verfahren als Möglichkeit zur ‚Manipulation‘, verzichten aber aus guten Gründen darauf, den Inhalt möglicher Manipulationen genauer festzulegen. Im Rahmen unserer Arbeit folgen wir vor allem der Annahme einer (zeit-) ökonomischen Orientierung der Interviewer. Als Manipulation in diesem Sinne gelten vor allem die Entscheidungen der Interviewer über den jeweils nächsten Kontaktversuch. Im Random Route-Verfahren bestehen weniger Möglichkeiten zur Kontrolle solcher Entscheidungen. Verzerrungen entstehen nach dieser engeren Deutung nicht in erster Linie dadurch, dass die Interviewer „das Interview mit der Person durchführen, die sie antreffen“ (Alt/Bien/Krebs 1991: 69), sondern dass sie bei mangelndem Erfolg im aktuellen Versuch – vielleicht entgegen der Regeln – häufig keinen weiteren Kontaktversuch mehr unternehmen. Natürlich können wir auch weitergehende Manipulationen nicht ausschließen.

2.2 Praktizierte Auswahlregeln auf der Ebene von Gebietseinheiten

In diesem Abschnitt werden alle ALLBUS-Erhebungen 1980–2000 nach dem Gesichtspunkt geordnet,

- in welchem Umfang den Interviewern innerhalb der Auswahlregeln Spielräume zugestanden wurden und
- in welchem Umfang Kontrollen der von ihnen durchgeführten Auswahl-schritte möglich waren. Offen bleibt dabei jedoch, inwieweit Kontrollen tatsächlich durchgeführt bzw. von den Interviewern der Institute erwartet wurden.

Die Beurteilung von Spielräumen und Kontrollmöglichkeiten stützt sich ausschließlich auf die Methoden-Berichte der ALLBUS-Erhebungen. Einzelheiten über alle 12 Erhebungen sind in schematischer Form in einem gesonderten Anhang wiedergegeben, der vom Autor bezogen werden kann. Hier beschränken wir uns auf die zusammenfassende Einordnung der Erhebungen in 4 Gruppen:

(1) ALLBUS-Erhebungen 1994/1996/2000

Den geringsten Spielraum sehen wir in den ALLBUS-Erhebungen mit einer Wahrscheinlichkeitsauswahl der Zielpersonen aus den Verzeichnissen der Einwohnermeldeämter in zuvor ausgewählter Gebietseinheiten. Innerhalb der Regeln ergeben sich hier nur Spielräume der Interviewer bei den Entscheidungen über die Wiederholung von Kontaktversuchen bei zunächst nicht erreichten Zielpersonen. Auch außerhalb der Regeln sind die Spielräume gering bzw. nur bei sehr ernsten Verstößen zu realisieren, wenn nämlich andere als die ausgewählten Personen befragt oder die Befragungen ganz gefälscht werden.

(2) ALLBUS-Erhebungen 1980/1982/1990/1998

Einen größeren Spielraum sehen wir in jenen ALLBUS-Erhebungen, bei denen die Auswahl der Haushalte vor der eigentlichen Erhebung erfolgte: Den Interviewern wurde eine feste Zahl von Haushaltsadressen zur möglichst vollständigen Realisierung je eines Interviews übergeben. Innerhalb der Haushalte folgte die Auswahl zwar festen Regeln (Schwedenschlüssel), war aber vom Interviewer selbst vorzunehmen und deshalb im Ergebnis dem Institut vorher nicht bekannt.

Unterschiede zu den Erhebungen der Gruppe (1) liegen deshalb aus unserer Sicht vor allem in der geringeren Kontrollierbarkeit der Auswahl zu befragender Personen. Bei strikter Einhaltung der Regeln wären beide Gruppen demgegenüber gleich einzustufen im Umfang des Spielraums bei der Auswahl: Es steht kein *Überschuss* an Adressen zur Verfügung. Die Entscheidungen der Interviewer über die erneute Kontaktierung zunächst nicht erreichter Haushalte oder – innerhalb derselben – beim ersten Kontakt mit dem Haushalt nicht anwesender Zielpersonen bestimmen den Grad der Ausschöpfung der Stichprobe.

(3) ALLBUS-Erhebungen 1984/1986/1988

Bei dieser Gruppe von Erhebungen wurde sowohl die Auswahl von Haushalten wie auch der zu befragenden Personen von den Interviewern vorgenommen. Auf einem vorgeschriebenen Weg durch das Erhebungsgebiet (random route) waren Haushalte in einer bestimmten Reihenfolge aufzunehmen und von diesen Haushalten eine bestimmte Zahl formal bestimmter Haushalte (z. B. auf bestimmten Positionen der

Liste von Haushalten, die vorher auf dem Formular gekennzeichneten waren) zu kontaktieren. Innerhalb der Haushalte erfolgte die Auswahl wieder aufgrund des Schwedenschlüssels.

Bei strikter Einhaltung der Regeln besteht auch in dieser Gruppe von Erhebungen weder ein Spielraum bei der Auswahl von Haushalten und Befragungspersonen noch ein Überschuss an potentiell zu befragenden Personen. Die Kontrollmöglichkeiten sind jedoch gegenüber den beiden erstgenannten Gruppen (1) und (2) noch weiter vermindert, da den Instituten zunächst weder die auszuwählenden Haushalte noch die darin anzusprechenden Befragungspersonen bekannt sind.

(4) ALLBUS-Erhebungen 1991/1992

Die Auswahl von Haushalten und darin Befragungspersonen folgt im Prinzip denselben formalen Regeln wie bei den Erhebungen der Gruppe (3) mit einer Ausnahme: Durch den festgelegten Zufallsweg (random route) im Erhebungsgebiet wird wieder nach bestimmten Regeln eine Liste von Haushalten mit einer Auswahl derselben erstellt. Diese Auswahl von Haushalten ist aber größer als die Zahl maximal zu realisierender Interviews. Wenn z. B. 6 Interviews in 15 ausgewählten Haushalten zu realisieren sind, so steht den Interviewern ein gewisser Überschuss an Haushalten zur Verfügung, aus denen Sie (innerhalb der Regeln!) frei wählen können. Dieses reduziert offensichtlich die Notwendigkeit – oder die wie immer sanktionierte Erwartung der Institute an ihre Interviewer – zur Wiederholung von Kontaktversuchen und spart damit Zeit und Geld.

Mit diesem Versuch der Rangordnung von ALLBUS-Erhebungen nach dem Grad der Spielräume, die Interviewern bei der Auswahl von Haushalten und/oder zu befragenden Personen offen stehen, verbinden wir entsprechend der oben formulierten Hypothese die Annahme von Folgen für die dabei entstehenden Wahrscheinlichkeitsauswahlen: *Schlecht erreichbare* Personen der Grundgesamtheit müssten gegenüber *gut erreichbaren* Personen um so häufiger ausfallen, je weniger sich Interviewer im Fall nicht erreichter Haushalte oder Befragungspersonen um die Wiederholung von Kontakten bemühen (müssen). Die Vermutung ist dabei, dass mit größeren Spielräumen – hier steigend von Gruppe (1) bis Gruppe (4) – auch der Druck zur Wiederholung von Kontaktversuchen abnimmt. Relativ schlechter erreichbare Personen müssten deshalb – ceteris paribus – gegenüber besser erreichbaren Personen besonders häufig in der Gruppe (4) und besonders selten in der Gruppe (1) ausfallen.

Aufgrund dieser Sichtweise richtet sich das Interesse also auf den unterschiedlichen Umfang systematischer Ausfälle in verschiedenen, von Wahrscheinlichkeitsauswahlen repräsentierten Teilpopulationen.

3 Die Untersuchung von Auswahlverzerrungen: Abweichungen von den erwarteten Verteilungen

In diesem Abschnitt wird ein Verfahren zur Beurteilung der Ausfallraten unterschiedlich gut erreichbarer Personen in Wahrscheinlichkeitsauswahlen beschrieben. Diesem Ziel liegt ein allgemeineres Problem zugrunde: Wie sind Ausfälle in Wahrscheinlichkeitsauswahlen zu analysieren, wenn man die Grundgesamtheit nicht vorher kennt?

Die Idee für eine – teilweise – Lösung dieses Problems haben wir an anderer Stelle beschrieben (Sodeur 1997). Die dabei verfolgte Lösungsstrategie ist auch nicht neu: Statt nach der offensichtlich schwierigen Lösung für ein Problem sucht man nach einer maßvollen Änderung des Problems mit dem Ziel, es dadurch lösbar zu machen und gleichzeitig wichtige Beiträge zur Lösung des ursprünglichen Problems zu leisten.

Dieser Strategie folgend wird anstelle der direkten Beurteilung einer Wahrscheinlichkeitsauswahl aus einer gegebenen Grundgesamtheit der Versuch unternommen, eine veränderte, das heißt in aller Regel eingeschränkte Grundgesamtheit zu definieren, über die man *infolge der getroffenen Einschränkung* wichtige Eigenschaften kennt. Diese Eigenschaften nutzt man anschließend zur Beurteilung der nachträglich auf gleiche Weise eingeschränkten Wahrscheinlichkeitsauswahl. Hilfreich wird dieses Verfahren vor allem dann, wenn mit der getroffenen Einschränkung von Grundgesamtheit und Wahrscheinlichkeitsauswahl keine für den Beurteilungszweck wesentlichen Eigenschaften ausgeschlossen werden.

In dieser Arbeit (wie in der Bezugsarbeit) definieren wir als neue, *eingeschränkte Grundgesamtheit* alle Personen, die mit Partnern des jeweils anderen Geschlechts im selben Haushalt zusammenwohnen, wobei auch die Partner zur Grundgesamtheit gehören müssen. Dieses ist nicht die einzig mögliche Form einer geeigneten Einschränkung der Grundgesamtheit. Sie hat aber den Vorteil, dass eine für viele inhaltlich begründete Fragen bedeutsame Grundgesamtheit übrig bleibt. Bei den ALLBUS-Erhebungen, auf die sich diese Untersuchung bezieht, bleiben damit (zunächst) rund zwei Drittel aller Personen der ursprünglichen Grundgesamtheit erhalten. Ausgeschlossen sind nur Befragte, die nicht mit einem festen Partner des anderen Geschlechts im selben Haushalt wohnen oder deren Partner nicht zur Grundgesamtheit gehört. Tabelle 1 gibt Aufschluss über die Fallzahlen der ursprünglichen Stichproben aller ALLBUS-Erhebungen von 1980 bis 2000 und über die Folgen der schrittweisen Eingrenzung auf eine Teilstichprobe mit den gewünschten Beschränkungen.

Tabelle 1 Abgrenzung der ausgewählten Teilstichprobe

ALLBUS Jahr	Stichprobe Gesamt West	Befragte ohne Partner	Partner nicht im Haushalt	Befragte mit Partner im Haushalt	Partner < 18 oder gleiches Geschlecht	untersuchte Teilstichprobe
1980	2955	1102	9	1844	65	1779
1982	2991	1178	0	1813	40	1773
1984	3004	1220	22	1762	26	1736
1986	3095	928	154	2013	7	2006
1988	3052	1056	134	1862	12	1850
1990	3051	984	195	1872	26	1846
1991	1514	469	77	968	5	963
1992	2400	723	139	1538	8	1530
1994	2342	612	118	1612	13	1599
1996	2402	565	147	1690	11	1679
1998	2212	758	89	1365	9	1356
2000	2036	526	142	1368	24	1344

Die auf diese Weise eingeschränkte Grundgesamtheit hat eine *definitorisch erzeugte Eigenschaft*: Sie besteht – wie jedes einzelne der sie konstituierenden, geschlechtsheterogenen Paare – aus genau 50% Frauen und 50% Männern. Gleiches gilt für alle Untermengen der Grundgesamtheit, soweit sie nicht aufgrund von Eigenschaften definiert werden, die mit dem Auswahlprozess zusammenhängen. Mit dem Erwartungswert eines Anteils von 50% Frauen an der Wahrscheinlichkeitsauswahl oder vielen Untermengen dieser Wahrscheinlichkeitsauswahl ist also ein allgemeines Kriterium zur Beurteilung der Auswahl bekannt.

Jedoch stellt eine derart abgegrenzte Wahrscheinlichkeitsauswahl nicht immer eine unverzerrte Auswahl der geschlechtsheterogenen Paare mit entsprechenden Eigenschaften dar, wie Günther Rösch (Büro für Erhebungsdesign und Datenanalyse, Frauenberg) kürzlich zu Recht in einem mündlichen Tagungsbeitrag kritisiert hat. Den meisten der hier analysierten ALLBUS-Erhebungen liegt eine mehrstufige Auswahl zugrunde (vgl. Abschnitt 2), auf deren letzter Stufe aus allen zur Grundgesamtheit gehörenden Haushaltsmitgliedern genau eines ausgewählt wird. Setzt sich ein Haushalt nur aus dem Paar und Kindern unter 18 Jahren (die nicht zur Grundgesamtheit gehören) zusammen, so wird auf der letzten Auswahlstufe eine von 2 Personen ausgewählt: Damit ist immer einer der beiden Partner dabei. In Haushalten mit drei zur Grundgesamtheit gehörenden Personen (z. B. neben dem Paar aus einem bereits volljährigen Kind oder einem Elternteil), so wird die zu befragende Person unter drei Personen gewählt und das Paar ist nur in zwei von drei

Fällen beteiligt. Entsprechendes gilt auch für größere Haushalte (immer bezogen nur auf die zur Grundgesamtheit gehörenden Personen): Bei 4 Personen im Haushalt fällt die Auswahl der zu befragenden Person in 2 von 4 Fällen, bei 5 Personen in 2 von 5 Fällen (usw.) auf einen der beiden Partner. Je größer also der Haushalt ist, mit umso geringerer Wahrscheinlichkeit wird ein Mitglied des betreffenden Paares für die Befragung ausgewählt und damit im Rahmen der oben beschriebenen eingeschränkten Wahrscheinlichkeitsauswahl berücksichtigt. Will man auch diese Verzerrung ausschließen, dann kann man die Analyse entweder unter Konstanthalten der Haushaltsgröße weiter beschränken, also hier auf Haushalte mit genau zwei zur Grundgesamtheit gehörenden Personen, nämlich den beiden Partnern. Oder man muss die Auswahlverzerrung in den größeren Haushalten durch ausgleichende Gewichtung korrigieren. Eine der beiden Korrekturen wird immer dann erforderlich sein, wenn der Verdacht besteht, dass die jeweils untersuchten Sachverhalte von der Haushaltsgröße zumindest teilweise mitbestimmt werden könnten. In der vorliegenden Arbeit verzichten wir auf Korrekturen dieser Art und beschränken uns auf den Hinweis, dass beide Versionen einer möglichen Korrektur nur zu sehr unwesentlich veränderten Ergebnissen führen.

Hinsichtlich der Haushaltsgröße folgen die Unterstichproben also im Prinzip Verzerrungen, die bereits in allen Gesamtstichproben mit Haushaltsbezug angelegt sind. Daneben gibt es aber auch andere Abweichungen zwischen der Gesamtstichprobe und der auf Paare bezogenen Teilstichprobe. Natürlich ist in der Teilstichprobe (nur Paare!) der Anteil der Verheirateten viel höher und der Anteil der Ledigen viel geringer als in der Gesamtstichprobe. Aber, etwas weniger offensichtlich, ist das Durchschnittsalter in der Teilstichprobe etwas höher (46,66 gegenüber 46,24 Jahre) und die Streuung etwas geringer als in der Gesamtstichprobe, was insbesondere an der geringeren Häufigkeit der jüngeren Jahrgänge unter den Befragten aus Paaren im gemeinsamen Haushalt liegt. Ebenfalls vor allem bei den jüngeren Personen haben unterschiedlich lange Ausbildungszeiten Auswirkungen auf die Paarbildung. So kommen in der Teilstichprobe wesentlich mehr Hauptschüler (56,5 gegenüber 50,1%) und weniger Personen mit höherem Schulabschluss (41,1 gegenüber 46,2%) vor als in der Gesamtstichprobe. Diese Unterschiede setzen sich fort bis zur jetzigen bzw. letzten beruflichen Stellung, wo Befragte aus den Paaren der Teilstichprobe seltener als Befragte aus der Gesamtstichprobe keine berufliche Position oder Ausbildungsposition nannten (5,7 gegenüber 10,7%).

Schließlich benötigen wir zur Prüfung der am Ende des vorangehenden Kapitels genannten Vermutungen nicht Erwartungswerte für den Anteil von Frauen oder Männern in der Wahrscheinlichkeitsauswahl, sondern entsprechende Erwar-

tungswerte für den Anteil relativ leicht gegenüber schwer erreichbarer Personen, um daran den Anteil der Ausfälle dieser Gruppen auf dem Weg von der Brutto- zur Nettostichprobe zu messen.

Ein entsprechendes Beurteilungskriterium gewinnt man durch Kombination unterschiedlicher Erreichbarkeiten mit dem – bei der hier gewählten Grundgesamtheit indikativen – Geschlecht der Befragten. Dazu muss man z. B. Paare auswählen, bei denen die Männer jeweils relativ schlecht und die Frauen relativ gut zu erreichen sind. Unter Anwendung von *Hilfstheorien* nimmt man z. B. Paare, bei denen der Mann berufstätig und die Frau nicht oder höchstens halbtags berufstätig ist. Die Unterschiede in den Erreichbarkeiten beider Partner lassen sich unter diesen Bedingungen verschärfen, wenn man die Auswahl der Paare zusätzlich einschränkt auf die Zugehörigkeit (kleiner) Kinder zum Haushalt und damit auf Anwesenheitsanfordernisse des nicht oder relativ wenig berufstätigen Partners.

Festzuhalten ist: die eben genannten Eigenschaften zur parallelen Beschränkung von Grundgesamtheit und Wahrscheinlichkeitsauswahl haben nichts mit dem Auswahlprozess zur Wahrscheinlichkeitsauswahl zu tun. Sie ändern deshalb auch nichts an dem Erwartungswert von 50% Frauen (bzw. Männern) in der Unterstichprobe. Die zusätzliche Beschränkung der Grundgesamtheit um bestimmte Konstellationen der Berufstätigkeit der jeweiligen Partner sowie um die Zugehörigkeit von Kindern zu ihrem Haushalt hat jedoch einen Preis (vgl. Kohler 2007: 59). Grundgesamtheit und zugehörige Wahrscheinlichkeitsauswahl werden dadurch weiter eingeschränkt, das zur Lösung ausgewählte Problem entfernt sich noch etwas mehr vom ursprünglichen Problem:

Angestrebt wird die Klärung der Zusammenhänge zwischen Auswahlverfahren, Erreichbarkeiten der Personen aus der Grundgesamtheit und ihrer tatsächlichen Aufnahme in die Befragung. Die Untersuchung beschränkt sich jedoch auf nur einen Teil der Grundgesamtheit, nämlich auf solche Paare im gleichen Haushalt, von denen die Männer aufgrund sehr spezieller Kriterien als relativ schlecht und die Frauen als relativ gut erreichbar eingestuft werden.

Definiert man z. B. eine Grundgesamtheit von geschlechtsheterogenen Paaren, bei denen der Mann ganztags berufstätig (relativ zur Frau schlecht erreichbar) ist und die Frau nicht oder höchstens halbtags berufstätig (gut erreichbar) ist und mindestens ein Kind bis zum Alter von 14 Jahren im Haushalt wohnt, so reduziert sich die Grundgesamtheit auf durchschnittlich 18,3% der ursprünglichen Größe⁶ oder in absoluten Zahlen: Von ursprünglich in den Allbus-Erhebungen der Jahre

6 Geschätzt anhand der entsprechenden Teilstichproben über alle 12 ALLBUS-Erhebungen 1980-2000 im Erhebungsgebiet West.

1980–2000 (im Erhebungsgebiet West) erfassten 31.054 Befragten gehören 19.461 zu den ausgewählten Paaren und darunter erfüllen 5.679 die genannten Bedingungen von Berufstätigkeit und Kindern.⁷

Im folgenden Abschnitt wird dieses Kriterium zur Prüfung der im Abschnitt 2 vermuteten, systematischen Auswahlverzerrungen benutzt. In einer Wahrscheinlichkeitsauswahl aus der eben beschriebenen Grundgesamtheit müssten Frauen und Männer ihre jeweiligen Paare mit gleicher Wahrscheinlichkeit *als Befragte* vertreten.

Abweichungen des Anteils befragter Frauen vom Erwartungswert (50%) indizieren also potentielle Verzerrungen der Auswahl, oder genauer: Aufgrund unterschiedlicher Erreichbarkeiten von Frauen und Männern in der oben definierten Grundgesamtheit müsste der Anteil der befragten Frauen einer verzerrten Wahrscheinlichkeitsauswahl systematisch und überzufällig höher als 50% liegen. Diese systematische Abweichung vom Erwartungswert müsste ferner mit dem Spielraum wachsen, der den Interviewern bei der Auswahl zugestanden und mutmaßlich in ihrem eigenen Interesse genutzt wird – entgegen den Regeln für die Erzeugung einer Wahrscheinlichkeitsauswahl.

4 Beurteilung der Ergebnisse

4.1 Vergleich der Rangordnungen nach Spielraum der Interviewer und Anteil befragter Frauen (leicht erreichbare Partner)

Tabelle 2 gibt Auskunft über den Umfang der Abweichungen im Anteil der befragten Frauen in den ausgewählten Untergruppen der 12 ALLBUS-Erhebungen. Spalte 2 enthält die Rangordnung der 12 Erhebungen nach dem Spielraum, der den Interviewern nach unserer Einschätzung bei der Wahrscheinlichkeitsauswahl der zu befragenden Personen zugestanden wurde (vgl. Abschnitt 2), Spalte 3 den Anteil der befragten Frauen (der Erwartungswert ist 50%!), und Spalte 5 die in z-Werte (d. h. mit dem Mittelwert 0 und der Standardabweichung 1) umgerechneten Anteile.

7 Beschränkt man sich zusätzlich aufgrund der oben diskutierten Auswahlverzerrungen bei unterschiedlichen Haushaltsgrößen auf Haushalte mit nur zwei erwachsenen Personen, so reduziert sich die Zahl der Paare weiter auf 5.040. Die folgenden Ergebnisse werden davon jedoch nicht nennenswert berührt.

Tabelle 2 Anteil befragter Frauen aus allen geschlechtsheterogenen Paaren, bei denen der Mann ganztags und die Frau höchstens halbtags berufstätig ist
(Nur Haushalte mit mindestens einem Kind unter 15 Jahren)

ALLBUS	Rang	% Frauen	N	z-Werte	Institut
1994	1	44,92	443	-2,1496	Infratest (Burke), München
1996	1	46,22	476	-1,6541	Infratest (Burke), München
2000	1	48,62	362	-0,5253	Infratest (Burke), München
1980	2	55,04	556	2,3890	GETAS, Bremen
1982	2	58,20	567	3,9587	GETAS, Bremen
1990	2	57,12	513	3,2585	INFAS, Bonn
1998	2	59,15	306	3,2562	GFM-GETAS (IPSOS), Hamburg
1984	3	54,16	517	1,8984	GETAS, Bremen
1986	3	55,97	620	2,9945	Infratest (Burke), München
1988	3	64,46	498	6,7419	GFM-GETAS (IPSOS), Hamburg
1991	4	60,82	319	3,9588	Infratest (Burke), München
1992	4	59,56	502	4,3644	Infratest (Burke), München

Die Erhebungen sind sortiert nach Einschätzung des Umfangs an Spielraum der Interviewer bei der Auswahl (Rang).

Sehr deutlich ist der Zusammenhang zwischen den Spielräumen der Interviewer und den Abweichungen des Frauen-Anteils in den Extremgruppen: Insbesondere die Erhebungen mit sehr geringem Spielraum (1), das sind die drei Erhebungen mit direkter Auswahl der Befragten aus den Listen der Einwohnermeldeämter, zeichnen sich durch Anteile von Frauen aus, die dicht beim bzw. sogar unterhalb des Erwartungswertes von 50% liegen.

Die Erhebungen mit dem vergleichsweise größten Spielraum (4) liegen entsprechend im oberen Bereich der Abweichungen mit einem besonders hohen Anteil an befragten Frauen. Sie werden in dieser Hinsicht nur vom ALLBUS 1988 übertroffen, dem nach unserer Einschätzung ein geringerer Spielraum (3) der Interviewer zugeschrieben wurde.

Weniger klar ist das Bild auf den beiden Stufen mittlerer Spielräume: Ohne Berücksichtigung des ALLBUS 1988 (Ausreißer) liegen die Abweichungen des Anteils befragter Frauen in der Gruppe (3) sogar noch unter den Anteilen in der Gruppe (2).

Insgesamt stimmen die Rangordnungen von Spielräumen der Interviewer und Anteilen der befragten Frauen jedoch recht gut überein: Kendall's tau-c= 0,65 ($p < 0,05$).

4.2 Diskussion der Abweichungen

Neben einer generellen Tendenz zur Bestätigung unserer Vorhersagen muss im Detail ein beträchtliches Maß an Abweichungen festgestellt werden. Einige dieser Abweichungen werden in diesem Abschnitt kommentiert und einige davon versuchsweise durch zusätzliche Eigenschaften der ALLBUS-Erhebungen erklärt. Im Gegensatz zur relativ groben Einstufung der Spielräume in 4 Rangstufen wurden diese zusätzlichen Eigenschaften erst nachträglich ermittelt.

(a) In allen drei Auswahlen auf der Basis der Einwohnermeldeamts-Listen liegt der Anteil der befragten Frauen unter dem Erwartungswert von 50%, zum Teil – im ALLBUS 1994 – mit mehr als 2 Standardabweichungen sogar sehr deutlich darunter. Dieses ist erstaunlich, weil durch unsere spezielle Konstruktion von Grundgesamtheit und Wahrscheinlichkeitsauswahl die Frauen systematisch mit der Eigenschaft relativ leichter Erreichbarkeit verbunden sind. Die Erwartung war deshalb generell auf einen Frauen-Anteil von über 50% gerichtet, und die Analyse zielte nur auf die Frage, um wieviel und in welchem Zusammenhang mit dem Spielraum der Interviewer der tatsächlich realisierte Anteil der Frauen über dem Erwartungswert lag.

Theoretische Erklärungen dieses unerwarteten Ergebnisses sind u. a., dass bei der direkten Auswahl der Personen über die Einwohnermeldeamts-Listen andere Auswahlverzerrungen wie etwa Rollenzuschreibungen innerhalb der Paare (wer nimmt an Interviews teil?) mit Folgen für Verweigerungen sichtbar werden, die unter anderen Auswahlbedingungen durch die starken, in entgegengesetzter Richtung wirkenden Erreichbarkeits-Effekte überlagert werden (diesen Hinweis verdanke ich einem mündlichen Hinweis meines Kollegen R. Schnell):

Die hier gefundene negative Abweichung des Frauen-Anteils vom Erwartungswert wäre damit (wie eine Abweichung in anderer Richtung auch) als Mischung der Folgen gegenläufiger Einflüsse zu interpretieren, nämlich der Folgen

- unterschiedlicher Erreichbarkeiten, die allein und ohne andere Einflüsse den Anteil der befragten Frauen innerhalb der systematisch ausgewählten Paare über den allgemeinen Erwartungswert von 50% heben müssten. Verzerrungen in dieser Richtung sollten bei Wahrscheinlichkeitsauswahlen auf der Basis von Einwohnermeldeamts-Listen zwar geringer sein als bei allen anderen Auswahlformen der ALLBUS-Erhebungen, aber gleichwohl bestehen bleiben
- aller anderen Einflüsse, die systematische Ausfälle überwiegend der Frauen, teilweise aber auch der Männer bewirken. Die oben genannten Vermutungen (Rainer Schnell) über die Ausfälle der Frauen würden wohl überwiegend den Verweigerungen zuzurechnen sein.

Zu den letztgenannten Einflüssen mag auch zählen, dass bei den Personenstichproben aufgrund der Ziehungen aus Einwohnermeldeamts-Listen Adressen und Zielpersonen vorab bekannt waren. Ein Teil der ersten Kontakte wurde deshalb vermutlich per Telefon geknüpft. Im Zusammenhang mit den ausgedehnten Nachbearbeitungsphasen aller drei Erhebungen mag es sein, dass die Männer trotz ihrer häufigeren Abwesenheit vom Haushalt über dieses Medium leichter zu erreichen oder zu einem Interview zu überreden waren.

Die stetige Abnahme dieser Abweichungen über die drei Erhebungen (1994–1996–2000) gab zunächst Anlass zur Vermutung, dass in irgendeiner Weise Korrekturmaßnahmen zur Vermeidung verzerrter Auswahlen ergriffen wurden, die später aufgrund zunehmender Erfahrungen mit Einwohnermeldeamts-Stichproben ihrerseits korrigiert wurden. Die Suche nach entsprechenden Hinweisen in den Methodenberichten führte zwar auf einige Unterschiede zwischen den Auswahlverfahren der drei Erhebungen, aber auf keine Stützung dieser Vermutung:

Die Unterschiede zwischen den drei Einwohnermeldeamts-Auswahlen betreffen einmal die Behandlung von Ersatzadressen für stichprobenneutrale Ausfälle, die 1994 unter etwas anderen Schichtungsbedingungen gezogen wurden als die ursprüngliche Bruttostichprobe (vgl. ALLBUS Methodenbericht, Koch/Wasmer/Harkness/Scholz 2001). In den Jahren 1996 und 2000 dagegen wurden diese Ersatzadressen unmittelbar zusammen mit der Bruttostichprobe (also nach gleichen Regeln) gezogen und den Interviewern zusammen mit den Adressen der zu realisierenden Befragungen übergeben. Die Ersatzadressen durften nur zum Ausgleich stichprobenneutraler Ausfälle eingesetzt werden.

Eine zweite Abweichung betrifft die Erhebung 2000 gegenüber 1994 und 1996: Bei sehr schlechter Ausschöpfung von Adressklumpen (0–2 Interviews) wurden diese vollständig durch neue Adressklumpen ersetzt, wobei auch die eventuell bereits durchgeführten Interviews der ersetzten Klumpen unberücksichtigt blieben. In beiden Fällen ist jedoch kein systematischer Zusammenhang mit dem geringen Anteil befragter Frauen auszumachen.

(b) Die Einstufung der Spielräume der Interviewer in die Klassen 2, 3 und 4 trägt nur wenig zur Ordnung der Anteile befragter Frauen bei, erklärt also nicht die Auswahlverzerrungen aufgrund unterschiedlicher Erreichbarkeiten der Partner im Sinne der oben genannten Hypothese. Insgesamt lässt sich die Varianz des Anteils befragter Frauen zu 80,4% (η^2) auf die Gruppierung in 4 Klassen von Spielräumen zurückführen. Fasst man demgegenüber die beiden mittleren Klassen (2) und (3) zusammen, so sinkt der erklärte Anteil kaum (80,1%), und gleiches gilt, wenn die Klassen (2)–(3)–(4) zusammengefasst werden (77,7%).

Als wirksam zur Erklärung von Ausfällen aufgrund unterschiedlicher Spielräume der Interviewer sind deshalb nur die beiden verbleibenden Klassen anzusehen:

- 2-stufige Auswahlen mit einer direkten Wahrscheinlichkeitsauswahl von Befragungspersonen aus den Einwohnermeldeamts-Listen gegenüber. (1)
- 3-stufige Auswahlen mit entweder Vorgabe ausgewählter Haushalte an die Interviewer oder Random Route-Verfahren zur Ermittlung der Haushalte durch die Interviewer auf der zweiten Stufe sowie einer Wahrscheinlichkeitsauswahl der Befragungspersonen innerhalb der Haushalte mittels Schwedenschlüssel durch die Interviewer auf der dritten Stufe. (2-3-4)

Dieses Ergebnis lässt sich auch nicht wesentlich dadurch verbessern, dass neben den bislang nur groben Einteilungen der ALLBUS-Erhebungen anhand weniger formaler Eigenschaften der Auswahlverfahren zusätzliche Kriterien zur Verfeinerung herangezogen werden. So wurden z. B. nicht in allen Erhebungen von den Interviewern Kontaktprotokolle geführt, die potentiell als Kontrollinstrumente eingesetzt werden könnten. Wo dieses geschah bleibt meist offen, welche Bedeutung die Institute der Dokumentation der Kontaktversuche und möglicher Ausfallgründe beigemessen haben und welche Konsequenzen daraus für Kontrollen und für die Steuerung des Interviewer-Einsatzes gezogen wurden. Eine wirklich verbesserte Einstufung der Spielräume wäre nur möglich, wenn detaillierte Informationen über die tatsächliche Vorgehensweise verschiedener Institute und – innerhalb gleicher Institute: über die Zeit – vergleichend zur Verfügung stünden.

Auf der anderen Seite haben wir vermutlich einigen formalen Kriterien zur Regelung des Auswahlverfahrens zu große Bedeutung beigemessen. So beruht z. B. die Abgrenzung zwischen den Klassen (3) und (4) auf folgenden Vorgaben an die Interviewer (hier nur typisiert wiedergegeben): In Erhebungen, die der Klasse (3) zugeordnet wurden, durften „maximal 10 Befragungen aus 10 Haushalts-Adressen“ realisiert werden, in Klasse (4) der gestuften Spielräume gab es die Erwartung von „6 Realisationen aus 15 Haushalts-Adressen“ (vgl. ALLBUS Methodenberichte).

Bei dieser Abgrenzung erschien die *explizite Freigabe von Spielräumen* an die Interviewer als entscheidender Gesichtspunkt. Demgegenüber muss vielleicht stärker betont werden, dass auch bei der scheinbar restriktiven Fassung „10 Realisationen aus 10 Adressen“ ein faktisch ganz ähnlicher, wenngleich nur implizit gewährter Spielraum bestand. Bei einer derzeit durchschnittlichen Ausschöpfung von rund 50% der Bruttostichprobe und einem entsprechend antizipierten Verhältnis von Brutto- und angestrebter Nettostichprobe ist jedem Interviewer klar, dass „5 Realisationen aus 10 Adressen“ keine schlechte Leistung bedeuten. Warum also sollten Befragungen in weiteren Haushalten versucht werden, wenn dieses erkenn-

bar, d. h. aufgrund zumindest eines vorangehenden Fehlversuches, nur mit größerem Aufwand erreichbar schien?

Zur realistischen Einschätzung der Situation von Interviewern müsste man also viel mehr wissen als nur die formalen Auswahlregeln:

- Welcher Art sind die Übereinkünfte zwischen Instituten und Interviewern über die Höhe der Bezahlung für realisierte Befragungen und welches die (übereinstimmenden?) Erwartungen über den Umfang der Nebenleistungen wie mehrfache Kontaktversuche, die mit der Bezahlung des Stücklohns als beglichen gelten?
- Wie werden mögliche Differenzen über solche Erwartungen zwischen Instituten und Interviewern abgestimmt?
- Gibt es über die formelle Dokumentation solcher Nebenleistungen (im Kontaktprotokoll) hinaus Prüfungen ihres Realitätsgehalts sowie tatsächliche oder von Interviewern erwartete Maßnahmen der Institute bei Verstößen?
- In welchem Umfang sind die eingesetzten Interviewer auf Beiträge in der laufenden Erhebung angewiesen bzw. stehen ihnen – möglicherweise günstigere – Befragungsalternativen offen?

Die Reihe von Fragen ließe sich leicht erweitern. Ihre Beantwortung verlangt Kenntnisse interner Verfahrensweisen der Institute bei der Feldorganisation, die für institutsübergreifende Vergleiche nicht leicht zu erlangen sein werden. Allerdings ließen sich mit solchen Erweiterungen der hier vorgeschlagenen Analyse wertvolle Einsichten in die Steuerung von Auswahlprozessen durch Interviewer gewinnen.

Literatur

- ADM Arbeitskreis Deutscher Markt und Sozialforschungsinstitute / AG.MA Arbeitsgemeinschaft Media-Analyse e.V. (Hg.), 1999: Stichproben-Verfahren in der Umfrageforschung. Opladen: Leske+Budrich.
- Alt, C., W. Bien und D. Krebs, 1991: Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtstichprobe. ZUMA-Nachrichten 28: 65-72.
- Bandilla, W., S. Gabler und M. Wiedenbeck, 1992: Methodenbericht zum DFG-Projekt ALLBUS Baseline-Studie 1991. ZUMA-Arbeitsbericht 92/04.
- Braun, M., R. Trometer und M. Wiedenbeck, 1989: Methodenbericht zur „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1988. ZUMA-Arbeitsbericht 89/02.
- Braun, M., C. Eilinghoff, S. Gabler und M. Wiedenbeck (1993): Methodenbericht zur „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1992.
- Brückner, E., H.-P. Kirschner, R. Porst, P. Prüfer und P. Schmidt, 1981: Methodenbericht zum „ALLBUS 1980“. ZUMA Arbeitsbericht 81/07.
- Erbslöh B. und M. Wiedenbeck, 1986: Methodenbericht zum „ALLBUS 1986“. ZUMA-Arbeitsbericht 87/04.
- Hagstotz, W., H.-P. Kirschner, R. Porst und P. Prüfer, 1982: Methodenbericht ALLBUS 1982. ZUMA-Arbeitsbericht 82/21.

- Kirschner, H.-P., 1984: ALLBUS 1980: Stichprobenplan und Gewichtung. S. 114-182 in: K. U. Mayer und P. Schmidt (Hg.): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Beiträge zu methodischen Problemen des ALLBUS 1980. Frankfurt, New York: Campus Verlag.
- Koch, A., S. Gabler und M. Braun, 1994: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1994. ZUMA-Arbeitsbericht 94/11.
- Koch, A., K. Kurz, H. Mahr-George und M. Wasmer, 1999: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1998. ZUMA-Arbeitsbericht 99/02.
- Koch, A., M. Wasmer, J. Harkness und E. Scholz, 2001: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS), 2000. ZUMA-Methodenbericht 01/05.
- Kohler, U., 2007: Surveys from inside: An assessment of unit nonresponse bias with internal criteria. *Survey Research Methods* 1: 55-67. <http://w4.ub.uni-konstanz.de/srm/>.
- Porst, R., P. Prüfer, M. Wiedenbeck und K. Zeifang, 1985: Methodenbericht zum "ALLBUS 1984". ZUMA Arbeitsbericht 85/03.
- Scheuch, E. K., 1974: Auswahlverfahren in der Sozialforschung. S. 1-96 in: R. König (Hg.): *Handbuch der empirischen Sozialforschung*, Band 3a, 3. Aufl. Stuttgart: Ferdinand Enke Verlag.
- Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske+Budrich.
- Sodeur, W. 1997: Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. *ZA-Information* 41: 58-82.
- Wasmer, M., A. Koch und M. Wiedenbeck, 1991: Methodenbericht zur "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 1990. ZUMA-Arbeitsbericht 91/13.
- Wasmer, M., A. Koch, J. Harkness und S. Gabler, 1996: Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 1996.
- Wasmer, M., A. Koch, 2002: Konzeption und Durchführung der PAPI-Methodenstudie zur "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2000. ZUMA-Methodenbericht 02/01.
- Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim, und Zentralarchiv für empirische Sozialforschung an der Universität zu Köln: Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS):
- Einzelstudie 1980, PC-lesbares Codebuch, ZA-Nummer 1000
 - Einzelstudie 1982, PC-lesbares Codebuch, ZA-Nummer 1160
 - Einzelstudie 1984, PC-lesbares Codebuch, ZA-Nummer 1340
 - Einzelstudie 1986, PC-lesbares Codebuch, ZA-Nummer 1500
 - Einzelstudie 1988, PC-lesbares Codebuch, ZA-Nummer 1670
 - Einzelstudie 1990, PC-lesbares Codebuch, ZA-Nummer 1800
 - Basisumfrage 1991, PC-lesbares Codebuch, ZA-Nummer 1990
 - Einzelstudie 1992, PC-lesbares Codebuch, ZA-Nummer 2140
 - Einzelstudie 1994, PC-lesbares Codebuch, ZA-Nummer 2400
 - Einzelstudie 1996, PC-lesbares Codebuch, ZA-Nummer 2800
 - Einzelstudie 1998, PC-lesbares Codebuch, ZA-Nummer 3000
 - Einzelstudie 2000, PC-lesbares Codebuch, ZA-Nummer 3451

Korrespondenzadresse: Wolfgang Sodeur
 Ostlandstraße 72
 50859 Köln
 wolfgang.sodeur@t-online.de

Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz

Empirische Befunde eines Methodenexperiments

Effects of prepaid monetary incentives on the return of mail survey and self-reporting about delinquency

Empirical results of an experiment

Rolf Becker, Regula Imhof und Guido Mehlkop

Zusammenfassung

Im Rahmen einer Studie über Determinanten delinquenten Handelns wird ein Methodenexperiment durchgeführt, um die Frage zu klären, ob auch bei einem sensiblen und heiklen Thema wie Kriminalität die Rücklaufquote über monetäre Anreize beeinflusst werden kann. Hierbei wird aus austausch- bzw. werterwartungstheoretischer Sicht die Annahme der strikten Rationalität zurückgewiesen, dass im Voraus beigelegte monetäre Anreize nicht zur Erhöhung des Rücklaufs beitragen. Gerade bei einer kompletten Befragung, in der Befragte über ihr delinquentes Handeln berichten sollen, können dem Fragebogen beigelegte Geldbeträge eine vertrauensbildende Maßnahme darstellen, die zu vergleichsweise hohen Rücklaufquoten führt. Hingegen wird angenommen, dass lediglich versprochenes Geld zu einem niedrigeren Rücklauf führt. Des Weiteren ist zu klären, ob monetäre Anreize zu einem

Abstract

In the context of an empirical study on determinants of delinquent behavior, a methodological experiment is carried out to investigate whether also for a sensitive topic like crime the return quota can be influenced by monetary incentives. The theoretical implication of strict rationality namely that prepaid monetary incentives do not contribute to the return quota positively in postal interviews is questioned from the theoretical perspective of the social exchange theory as well as from the theory of subjectively expected utilities. In a survey in which people are asked to report about their former delinquent behavior and, in particular, about delinquency in the future, prepaid monetary incentives might be a confidence-building measure which results in comparatively high return quotas. Merely promising money might lead to a lower return since this strategy contributes to respondents' non-participation. Our em-

‚Overreporting‘ der Delinquenz führt. Die empirischen Befunde bestätigen unsere Annahmen eindeutig, auch wenn die sozialen Mechanismen der Reziprozität nicht direkt beobachtet werden. Ebenso zeigt sich, dass monetäre Anreize weder soziale Erwünschtheit noch ‚sponsorship effects‘ auslösen, die zu einem ‚Overreporting‘ von bereits begangenen oder antizipierten Straftaten führen. Schließlich bestätigte sich für sozialstrukturelle Merkmale und Korrelate des delinquenten Handelns, dass sich die drei Teilstichproben nicht systematisch voneinander unterscheiden.

empirical findings confirm these assumptions. However, due to the lack of further information it remains unsolved, whether reciprocity and trust are actually the main mechanisms important for filling out and returning the questionnaire. Furthermore, we find that monetary incentives do not contribute to overreporting delinquency due to social desirability. Finally, there are indications that the three partial samples do not differ significantly with regard to social background and other characteristics correlating with delinquency.

1 Einleitung¹

Bei Befragungen, in denen die Befragten selbst über ihr gesetzeswidriges und damit strafbares Handeln berichten sollen, tritt zum einen das methodische Problem auf, die Befragten generell zur Teilnahme an der (freiwilligen) Befragung zu bewegen. Zum anderen muss bei Teilnehmern sichergestellt werden, dass sie valide wie reliable Antworten liefern (Hindelang/Hirshi/Weis 1979). Bei solchen, für die Befragten heiklen wie unangenehmen Fragen, müssen deswegen besondere Vorkehrungen getroffen werden, um die Anonymität der Angaben zu garantieren. Solche Vorkehrungen sind besonders bei Fragen zu rechtswidrigem Verhalten notwendig, wenn die Respondenten negative Sanktionen erwarten (Porst 2001; Mehlkop/Becker 2007). Im ALLBUS 1990 und 2000 wurde beispielsweise die so genannte ‚sealed envelope technique‘ angewandt, die als ein geeignetes Vorgehen anzusehen ist, Befragten absolute Anonymität und Vertraulichkeit bei Fragen zu deren Delinquenz zu garantieren (Becker 2006; de Leeuw 2001; Sudman/Bradburn 1974, 1982). Dadurch sollten sowohl verzerrte Antworten als auch Antwortverweigerungen unwahrscheinlicher werden. Für den ALLBUS 2000 konnte festgestellt werden, dass auch bei Verwendung eines vertraulichen, selbst auszufüllenden Fragebogens die

1 Für hilfreiche Kommentare danken wir M. Abraham, R. Porst und B. Jann. Für Hilfestellungen bei Datenerfassung und Datenaufbereitung danken wir G. Ludi, M. Raimann und S. Zingaro. Weiterhin bedanken wir uns beim Schweizerischen Nationalfonds (SNF) für die finanzielle Förderung des Forschungsprojektes „Kriminalität in Bern“ im Allgemeinen und des Methodenexperiments im Besonderen. Schließlich sind wir den anonymen Gutachtern und der Redaktion zu Dank verpflichtet für ihre hilfreichen Kommentare.

Merkmale der Interviewer und der Interviewersituation die Tendenzen des Befragten zu Antwortverweigerungen oder zu ‚wahren‘ Antworten zum delinquenten Handeln beeinflussen (Becker/Günther 2004).

Aufgrund dieser Interview- und Interviewereffekte wird die *postalische Befragung (mail survey)* als besonders geeignete Befragungsmethode favorisiert (Sudman/Bradburn 1982). Denn bei postalischen Befragungen, die den Befragten oftmals als weniger aufdringlich erscheinen als andere Befragungsmethoden, sind Anonymitätsgarantien eindeutig und Interviewereffekte ausgeschlossen (Klein/Port 2001). Dieser Vorschlag wird durch eine Vielzahl methodischer Studien empirisch gestützt. So zeigen beispielsweise Sudman/Bradburn (1982: 42-44), dass bei Fragen zu rechtswidrigem Verhalten die postalische Befragung eher ‚wahre‘ Antworten erzielt als andere Befragungsmethoden. Soziale Erwünschtheitseffekte bei heiklen Fragen sind bei postalischer Befragung generell seltener (Krysan et al. 1994). Auch ist dann bei heiklen Fragen die Datenqualität besser:

„In general, it is somewhat harder to have people answer questions in the mail survey as the higher items missing data indicates, but when the questions are answered, the resulting data are of better quality (more self-disclosure, more precision)“ (de Leeuw 1992: 77).

Diese Erkenntnisse haben uns dazu bewogen, im Vorfeld einer *postalischen Befragung über kriminelles Handeln* von erwachsenen Einwohnern in Bern, ein Methodenexperiment durchzuführen. Empirisch festgestellt werden soll, ob monetäre Anreize den Rücklauf bei einer *kompletten Befragung zur eigenen Delinquenz* in bedeutsamer Weise fördern. Zudem ist zu klären, ob monetäre Anreize auch die Qualität der Antworten auf Fragen zur Neigung für Delinquenz und Prävalenz strafbarer Handlungen in besonderer Weise beeinflussen.

Wie Berger (2006: 95) zu Recht in seinem Literaturbericht feststellt, gibt es im deutschen Sprachraum (noch zu) wenige Erfahrungen mit unkonkonditional beim Erstkontakt gegebenen Geldbeträgen. Somit ist – anknüpfend an ein früheres Methodenexperiment mit einem anderen Design (Mehlkop/Becker 2007) – die vorliegende Untersuchung ein weiterer Beitrag zur Schließung dieser Lücke. Im folgenden Abschnitt werden in aller Kürze Forschungsstand im Lichte empirischer Ergebnisse und theoretische Erklärungsansätze zur Effektivität monetärer Anreize bei postalischen Befragungen diskutiert. Die Darstellung unserer Studie über kriminelles Handeln, der Daten und des experimentellen Designs bildet den dritten Abschnitt. Die empirischen Resultate werden im vierten Abschnitt dargestellt und eine abschließende Diskussion erfolgt im fünften und letzten Abschnitt.

2 Forschungsstand und theoretischer Hintergrund

2.1 Forschungsstand im Lichte empirischer Ergebnisse

Betrachtet man sowohl Einzelstudien als auch Meta-Analysen mehrerer Studien, dann liegen vor allem für die aktuellen Experimente eindeutige Befunde und einhellige Interpretationen der Ergebnisse vor (Singer 1998): Monetäre Anreize führen bei postalischen Befragungen zu höheren Rücklaufquoten als in den Fällen, in denen den Befragten kein Geld zum Fragebogen beigelegt wird (Berger 2006; Jobber et al. 2004: 21; Harkness et al. 1998; Church 1993; Yu/Cooper 1983; Armstrong 1975: 111; Kanuk/Berenson 1975). Aber es gibt offensichtlich keinen strikten linearen Zusammenhang zwischen der Höhe des Geldgeschenkes und dem Rücklauf in dem Sinne, dass die Antwortbereitschaft unbegrenzt mit steigendem Geldbetrag zunimmt (Linsky 1975: 96-98; Heberlein/Baumgartner 1978; Goyder 1982; Yu/Cooper 1983; Yammarino/Skinner/Childers 1991; Singer et al. 1998; Berger 2006: 89-90). Entscheidend für die Steigerung des ‚Responses‘ ist, dass den Befragten beim Erstkontakt das Geld *sofort und bedingungslos* gemeinsam mit dem Fragebogen – als Gabe im Voraus – zukommt (Church 1993: 75; James/Bolstein 1992: 442-443; Furse/Stewart 1981; Armstrong 1975: 116). Werden den Respondenten hingegen lediglich Geldbeträge in Aussicht gestellt, sofern sie den Fragebogen ausgefüllt zurückschicken, so sind die Rücklaufquoten oftmals signifikant niedriger, als wenn Respondenten den Geldbetrag bereits im Voraus gemeinsam mit dem Fragebogen erhalten (Church 1993; Diekmann/Jann 2001). Offensichtlich bieten *versprochene* Geldbeträge – unabhängig von ihrer Höhe – keinen signifikanten zusätzlichen Anreiz für die Befragungspersonen, den Fragebogen auszufüllen und zurückzusenden (James/Bolstein 1992: 451; Schewe/Gournoyer 1976: 105; Armstrong 1975: 113). Eher weisen in der Zwischenzeit publizierte Ergebnisse darauf hin, dass Rücklaufquoten bei versprochenen Geldbeträgen tendenziell niedriger sind als wenn Befragungspersonen kein materieller Anreiz angeboten wird. Ebenso wenig effektiv ist es, den Befragten Teilnahmen an Lotterien oder die Möglichkeit, dass die Forscher in ihrem Namen für wohltätige Zwecke spenden, anzubieten (Stadtmüller/Porst 2005: 7; Warriner et al. 1996; Hubbard/Little 1988; Furse/Stewart 1982).

Befürchtungen, dass durch monetäre Vorleistungen systematische Verzerrungen in der realisierten Stichprobe entstehen oder die Qualität von Antworten (z. B. wegen sozialer Erwünschtheit oder Sponsorship-Effekte) und damit die Datenqualität beeinflusst werden können, scheinen auf den ersten Blick plausibel zu

sein (Stadtmüller/Porst 2005: 10–11).² Jedoch sind empirische Befunde dazu uneinheitlich (Berger 2006: 93–95). So weisen James/Bolstein (1992) nach, dass ‚prepaid monetary incentives‘ geringe Effekte auf die sozialstrukturelle Zusammensetzung der Gruppe der Teilnehmer haben (Arzheimer/Klein 1998). Nederhof (1983) hingegen stellt in seiner Studie fest, dass gerade untere Bildungs- und Sozialklassen überproportional erreicht werden. Zu anderen Befunden kommt zuvor Gelb (1975): Die Gabe von 50 Cent (in US-Währung) resultierte in höheren Rücklaufquoten aus den Mittelschichten, während die unteren Sozialschichten eher auf versprochene Geldbeträge reagierten. Der Autorin zufolge scheint es so zu sein, dass sich die Klassendifferenzen im Rücklauf weiter erhöhen, wenn eine sofortige Belohnung mit steigenden Geldbeträgen erfolgt.

Des Weiteren stellt Nederhof (1983) fest, dass die Antworten der Befragten nicht durch die materiellen Anreize verzerrt werden; eher wird die Qualität der Angaben erhöht, da die ‚Incentives‘ die Befragten zu zusätzlichen Anstrengungen bei der Beteiligung an der postalischen Befragung und zu kompletten Antworten anreizen. Zu ähnlichen Befunden gelangen McDaniel und Rao (1980): Monetäre Geschenke erhöhen nicht nur den Rücklauf, sondern auch die Qualität der Daten infolge genauerer und vollständiger Angaben durch die Befragten (vgl. Berger 2006: 96). Diese Befunde werden durch Darven et al. (2003: 140) bekräftigt: Monetäre Anreize im Voraus erhöhen gleichermaßen den ‚Response‘ wie die Wahrscheinlichkeit für komplette und konsistente Antworten (vgl. James/Bolstein 1990).

2.2 Theoretische Erklärungsansätze

Wie kann man erklären, dass monetäre Anreize, die als Vorleistungen den Befragten übergeben werden, zu vergleichsweise höheren Rücklaufquoten führen? Die Effektivität dieser Strategie und Maßnahme, die Rücklaufquote bei postalischen Befragungen zu erhöhen, wird in der Regel anhand *entscheidungs- und handlungstheoretischer Ansätze* erklärt: „Ausgangspunkt aller Überlegungen ist die Annahme, dass die Entscheidung zur Teilnahme an einer (schriftlichen) Befragung nach Abwägung von Vorteilen (Nutzen) und Nachteilen (Kosten und Mühen) durch die zu befragende Person erfolgt: Nur wenn der erwartete Nutzen der Teilnahme die er-

2 Denn mit einer hohen Ausschöpfungsquote ist die Datenqualität nicht von vornherein garantiert (Koch 1998). So ist nicht ausgeschlossen, dass die Bereitschaft, auf Fragen zur eigenen Delinquenz zu antworten, wegen monetärer Anreize mit individuellen Dispositionen und sozialstrukturellen Merkmalen korreliert. Vor allem im Falle, dass die erhöhte Antwortbereitschaft mit Fragen der eigenen Delinquenz korreliert, erhält man ebenfalls eine erheblich verzerrte Stichprobe, die bei multivariaten Analysen zu irreführenden Ergebnissen über delinquentes Handeln führen kann.

warteten Kosten überwiegt, sind Personen bereit, sich an einer Umfrage zu beteiligen" (Porst 2001: 1; Esser 1974, 1986).³ So können bei einer postalischen Befragung Forscher dazu beitragen, die potentielle Befragungsperson zum Ausfüllen und Rücksenden eines Fragebogens zu motivieren, indem sie versuchen, die Kosten für die potentielle Befragungsperson zu minimieren und den Nutzen zu maximieren. Vor allem von Dillman (2000) wird wiederholt darauf hingewiesen, dass zusätzlich zur intrinsischen Motivation von Befragungspersonen der erwartete Nutzen durch *extrinsische Anreize in Form monetärer Geschenke* erhöht werden kann. Gerade bei postalischen Befragungen, in denen die Befragten einen (teilweise recht umfangreichen) Fragebogen allein ohne einen motivierenden und beim Ausfüllen des Fragebogens behilflichen Interviewer ausfüllen müssen, scheinen extrinsische Anreize nötig, um von Befragten subjektiv wahrgenommene Kosten zu senken (Mehlkop/Becker 2007; Hill/Willis 2001: 416; Singer 1998: 9).

Des Weiteren ist aus Gründen der ‚*bounded rationality*‘ (Simon 1993) anzunehmen, dass viele Menschen in einer Befragungssituation unentschlossen sind, da sie aufgrund fehlender oder unvollständiger Informationen entstehende Kosten und Nutzen nicht eindeutig diskontieren können. Gerade bei dem Thema ‚Kriminalität‘ sollte man davon ausgehen, dass die Gruppe der Unentschlossenen recht groß ist, weil die Befragten sich darüber unsicher sind, ob sie Informationen über eigenes kriminelles Handeln preisgeben sollen (Mehlkop/Becker 2007). So sind die extrinsische Motivation fördernde monetäre Geschenke zum einen deswegen wichtig, weil bei einer postalischen Befragung die Antwortverweigerung den Befragten leichter fällt als bei Anwesenheit eines Interviewers. Somit muss die Befragungsperson einen Nutzen für sich sehen, der durch zusätzliche Anreize garantiert scheint. Da zum anderen bei postalischen Befragungen in der Regel die Anonymität garantiert ist und wegen Abwesenheit des Interviewers auch keine unmittelbaren Sanktionen zu befürchten sind, steht dann eher die Kompensation anfallender Partizipationskosten zu Gunsten höherer Teilnahmebereitschaft im Vordergrund. So können

3 Der Nutzen kann beispielsweise soziale Anerkennung als kooperative Befragungsperson, Erfüllung einer Höflichkeitsnorm gegenüber Fremden, Interesse am Thema der Befragung, Mitteilung eigener Meinungen, Loyalität zu einer öffentlichen Einrichtung oder Unterstützung der Wissenschaft sein. Kosten hingegen ergeben sich unter anderem aus Opportunitätskosten der Zeit, kognitiven Leistungen, Preisgabe privater Informationen, Bekanntgabe ‚wahrer‘ Angaben bei heiklen Fragen oder Befürchtung von Sanktionen (Becker 2006). „Andererseits kann die Antwortverweigerung dann Nutzen erbringen, wenn beispielsweise Konformität mit Normen von bestimmten (delinquenten) Bezugsgruppen oder grundsätzliche Abneigung gegenüber wissenschaftlichen Befragungen erwartet wird. Kosten sind beispielsweise mit der subjektiv erwarteten Missbilligung von bestimmten Antworten durch den Interviewer verbunden. Des Weiteren führt Esser (1986) Transaktionskosten an, die für die befragte Person durch ihre Unsicherheit im Umgang mit einer fremdartigen Situation wie dem Interview und durch Verständnisprobleme während des Interviews entstehen können“ (Becker/Günther 2004: 42-43).

prinzipiell drei Gruppen von Befragten unterschieden werden: Die erste Gruppe sind Befragte, deren intrinsische Motivation allein zur Teilnahme führt. Die zweite Gruppe besteht aus Befragten, die grundsätzlich nicht zur Teilnahme an der Befragung bereit sind und die auch nicht durch monetäre Anreize dazu gebracht werden können. Die dritte (und vermutlich zahlenmäßig recht große Gruppe) besteht aus unentschlossenen Befragungspersonen. Gerade in dieser letzten Gruppe sollten monetäre Anreize eine entscheidende Rolle spielen, denn sie scheinen besonders dann eine Rücklaufsteigerung zu bewirken und vormals Unentschlossene (oder auch Ablehnende) zur Mitwirkung motivieren zu können, wenn sie im Voraus, und ohne an Bedingungen geknüpft zu sein, gegeben werden (Mehlkop/Becker 2007). Somit können die Kosten kompensierende Anreize die individuelle Kosten-Nutzen-Bilanzierung zu Gunsten des Nutzens beeinflussen, wenn die erwarteten Kosten einer Teilnahme durch monetäre Gaben ausgeglichen werden (Hill/Willis 2001: 416; Riphahn/Serfling 2002: 8). Gerade die Gruppe der wegen ungünstiger bzw. unklarer Kosten-Nutzen-Bilanzierung unentschlossenen Befragungspersonen kann dadurch zum sicheren Entschluss bewegt werden, doch noch an der Befragung teilzunehmen (Schräpler 2001: 5).

Der strikten Variante von Rationalität zufolge wäre allerdings zu erwarten, dass sich bei ‚prepaid monetary incentives‘ die Rücklaufquote ungefähr dem Rücklauf ohne monetäre Anreize entspricht, da die Befragungspersonen das Geld ohne Beantworten des Fragebogens als Gegenleistung einstecken würden (Stadtmüller/Portst 2005: 4; Diekmann/Jann 2001).⁴ Auch wenn die monetären Anreize zwar die Kosten-Nutzen-Abwägung der angeschriebenen Person beeinflussen, so müssen zum einen die Befragungspersonen bei anonymen Befragungen keine Konsequenzen erwarten, wenn sie das Geld ohne Gegenleistung behalten, weil zum anderen die Geldgeschenke kaum die Kosten der Teilnahme kompensieren und eher

4 Der restriktiven Rational Choice Theorie zufolge verhalten sich Individuen im Sinne einer perfekten Rationalität in jeder Entscheidungssituation in optimaler Weise nutzenmaximierend, da ihnen alle Informationen zu alternativen Handlungen und deren Konsequenzen bekannt sind. Hingegen kann aus der Perspektive der subjektiven Werterwartung davon ausgegangen werden, dass Personen aus einem subjektiv wahrgenommenen Set von Handlungsalternativen diejenige auswählen, von der sie sich den größtmöglichen subjektiven Nutzen erwarten. Da Menschen fehlbar, vergesslich und zuweilen auch bequem sind und nur über begrenzte Ressourcen (z. B. Zeit, kognitive Möglichkeiten, Gelegenheiten, etc.) zur Informationssuche und der Verarbeitung wahrgenommener Informationen verfügen, ist statt von einer perfekten eben von einer eingeschränkten Rationalität auszugehen. Individuen verwenden für die Analyse und Bewertung der sozialen Situation, Handlungsalternativen und Konsequenzen einfache kognitive Modelle und folgen diesen bei der Auswahl von Handlungen. Aufgrund dieser kognitiven und sozialen Restriktionen erfolgt die Wahrnehmung, Evaluation, Entscheidung und Handlung nach Kosten-Nutzen-Prinzipien, die für das Individuum zufrieden stellend sind und zu suboptimalen statt perfekt optimalen Ergebnissen führen.

als zusätzlicher Nutzen definiert werden. Während in diesem Falle die Strategie ineffektiv ist, um den Rücklauf zu erhöhen, wäre das Versprechen von Geld als Gegenleistung für die Beantwortung des Fragebogens erfolgreicher: „Die Hypothese strikter Rationalität prognostiziert, dass versprochene, nicht aber beigelegte Geschenke die Ausschöpfungsquote erhöhen werden“ (Diekmann/Jann 2001: 21). Aber die empirischen Befunde sprechen – wie bereits gesehen – gegen diese Hypothesen strikter Rationalität der Befragungspersonen.

Ein Ausweg aus dem Erklärungsdilemma scheint die auf der Austauschtheorie beruhende *Reziprozitätsthese* zu sein (Dillmann 2000), die – weil die Austauschtheorie ebenfalls auf der Prämisse individueller Kosten-Nutzen-Abwägung aufbaut (Homans 1961) – ohne größere Schwierigkeiten in eine aufgeklärte Variante der Rational Choice Theorie wie etwa der Theorie subjektiver Werterwartung integriert werden kann. Eine (schriftliche) Befragung wird als sozialer Austausch verstanden (Esser 1986), der dann zustande kommt, wenn beide Parteien – der Sozialforscher und die Befragungsperson – sowohl einen gewichtigen Grund zur Annahme haben, davon zu profitieren, als auch unterschiedliche Präferenzen und Bedürfnisse haben: Information gegen interessante Fragen. Der Unterschied in der Informiertheit beider Interaktionspartner ist Motivation für die befragte Person, den Fragebogen zu beantworten. Das an keine Bedingungen geknüpfte Geldgeschenk erfordert Nachgiebigkeit bei der Beantwortung gestellter Fragen. Demnach kann der monetäre Anreiz nicht den Nutzen aus der Teilnahme an der Befragung herstellen, da die Befragungsperson das Geld ohne Gegenleistung behalten kann: „Das Incentive eröffnet vielmehr eine soziale Austauschbeziehung mit dem Respondenten: Zunächst ohne Gegenleistung überreicht, fungiert es als ein ‚symbol of trust‘ (Dillman 2000) und generiert ein Gefühl der sozialen Verpflichtung des Befragten, dem Wunsch nach einem Interview zu entsprechen. Die Interviewteilnahme stiftet so durch die Erfüllung der so genannten Reziprozitätsnorm dem Befragten einen zusätzlichen Nutzen. Die Reziprozitätsnorm besagt, dass man bemüht sein sollte, freiwillig gewährte Vergünstigungen in irgendeiner Form ‚zurückzuzahlen‘ (Gouldner 1960; Mauss 1984: 157). Der Wert des Incentives sollte vor diesem Hintergrund deshalb auch gegenüber den Befragten nicht als ‚Vergütung‘ für die Interviewteilnahme dargestellt werden. (Arzheimer/Klein 1998: 8; vgl. Groves/Singer/Corning 2000).⁵

5 Insbesondere wenn solche Geschenke als Gabe im Voraus übergeben werden, fühlen sich die ‚beschenkten‘ Befragten verpflichtet, mit der Teilnahme an der Befragung und Beantwortung der (auch für sie heiklen) Fragen eine adäquate Gegenleistung zu liefern. Dann sollte – wie bereits zuvor begründet – der aus der Teilnahme an der Befragung resultierende Vorteil die antizipierten Kosten deutlich übersteigen, und die vormalis Unentschlossenen werden deswegen zur Teilnahme motiviert.

Allerdings hätte die Befolgung der Reziprozitätsnorm die drohende Sanktionierung bei Nichtbefolgung als Voraussetzung; aber bei sozialwissenschaftlichen Befragungen ist die Teilnahme der Respondenten freiwillig und die Forscher verfügen über keine Sanktionsmöglichkeiten. So wird entsprechend der Austauschtheorie von Blau (1964) die Gegenseitigkeit nicht durch soziale Normen erzwungen, sondern soziale Reziprozität als Ergebnis der Bedingungen, die dem Austauschprozess innewohnen, kann auch ohne normative Kontrollen bestehen. Denn Reziprozität entsteht durch das Interesse an dem Nutzen des sozialen Austauschs, der sich durch die Erfüllung von Verpflichtungen ergibt. Im Sinne der Überlegungen von Blau (1964) begibt sich aber eine Befragungsperson, die eine Gabe des Sozialforschers nicht erwidert und somit einer Verpflichtung nicht nachkommt, in ein Abhängigkeitsverhältnis. Um eine Abhängigkeit zu verhindern, zumal diese eher mit Nachteilen als mit Vorteilen verbunden ist, versucht diese Person eine erwartete wie angemessene Gegenleistung zu erbringen und damit ihre Verpflichtungen zu erfüllen.

Da bei ‚mail surveys‘ keine Rückzahlungen im Vorfeld der Befragung vereinbart werden, verlangt eine Befragung als sozialer Austausch ein gewisses Maß an Vertrauen in sozialen Beziehungen zwischen Interviewer und Interviewten. Zu Anfang dieser Beziehung ist nur ein Minimum an Vertrauen notwendig, weil die Anfangsinvestition ein geringes Risiko birgt. Die implizite Forderung, sich als kooperative Befragungsperson erkenntlich zu zeigen, um weiteren Nutzen aus dem Austausch zu erlangen, dient als ein ‚starting mechanism‘ (Gouldner 1960). Wenn Sozialforscher und Befragungsperson in Kontakt treten, bevor sich gemeinsame Normen, Ziele oder Erwartungen an das Interview als Austauschprozess herauskristallisiert haben, liefern die aus solchen Tauschbeziehungen resultierenden Vorteile auch die Motivation für die Befragung. Im Sinne der impliziten Rational Choice Theorie kann – wie bereits angesprochen – der Nutzen aus der Teilnahme an einer wissenschaftlichen Befragung aus dem Wunsch nach Artikulation der eigenen Meinung und Identität, der Unterstützung von Wissenschaft, dem Interesse am Thema der Befragung oder an sozialer Anerkennung als ‚gute‘ Befragungsperson, resultieren (Esser 1986, 1974).

2.3 Hypothesen

Aus den theoretischen Ausführungen lassen sich folgende Hypothesen ableiten, die zu ‚erklären‘ vermögen, weshalb Befragungspersonen über monetäre Anreize zur Teilnahme an postalischen Befragungen motiviert werden können. Offensichtlich erzeugt das dem Fragebogen beigelegte Geld neben dem Gefühl der Reziprozität

auch Vertrauen bei den angeschriebenen Befragungspersonen gegenüber den Forschern bzw. der Studie: „Sending the money with the questionnaire is a goodwill gesture that puts the sponsor and questionnaire in a positive light and sets the stage for the respondent to reciprocate with an appropriate gesture of completing the questionnaire“ (Dillman 2000: 168). Allerdings darf der Geldbetrag nicht allzu hoch sein (z. B. dem tatsächlichen Gegenwert der von der Befragungsperson zu erbringenden Leistung entsprechen), da dann ein auf Vertrauen basierendes Reziprozitätsverhältnis von den Befragungspersonen als geschäftsmäßiger Austausch definiert werden kann: „Unter diesem Gesichtspunkt wird die potentielle Befragungsperson dann prüfen, ob diese ‚Bezahlung‘ eine adäquate Vergütung für ihre Mühen darstellt“ (Stadtmüller/Porst 2005: 5).⁶

(1) Im Unterschied zur restriktiven Hypothese strikter Rationalität wird aus austausch- und werterwartungstheoretischer Sicht erwartet, dass nicht allzu niedrige und nicht allzu hohe monetäre Anreize zur Erhöhung der Ausschöpfungs- und Rücklaufquoten von postalischen Befragungen – auch zu sensiblen und für die Befragten heiklen oder peinlichen Fragen zur eigenen Delinquenz – beitragen, da sie dann eher den Eindruck gewinnen, dass sich die Teilnahme lohnt bzw. ihnen wenigstens nicht schadet (vgl. Schnell 1997: 251). Bei monetären Leistungen im Voraus werden die angeschriebenen Befragungspersonen eher antworten als diejenigen, die keine monetären Anreize erhalten haben.

(2) Ebenfalls wird im Unterschied zur restriktiven Hypothese strikter Rationalität angenommen, dass ein versprochenes Geschenk *nicht* zur signifikanten Erhöhung der Rücklaufquoten postalischer Befragungen führt. Im Gegenteil: Da ein versprochenes, an Vorleistungen geknüpftes Geschenk möglicherweise eher als ein ‚*symbol of distrust*‘ oder als der Beginn eines geschäftsmäßigen Austauschs (d. h.

6 US-amerikanische Studien und eine Vielzahl von Meta-Analysen berichten über einen kurvilinearen Zusammenhang zwischen Geldbetrag und Rücklauf (Armstrong 1975; Fox/Crask/Kim 1988; James/Bolstein 1992; Warriner et al 1996; Berger 2006). Zum einen variieren die Geldbeträge als Schwellenwert für zu hohe Vorleistungen und zum anderen unterscheiden sich die jeweiligen Studien in Thema und Design. So ist beispielsweise nicht sichergestellt, ob der Schwellenwert für eine Studie zu einem unverfänglichen Thema einen gültigen Schwellenwert für Befragungen zu heiklen Themen darstellt. Diese offene Frage eröffnet zweifelsohne Perspektiven für zukünftige Methodenforschung, in der die Geldbeträge variiert werden. Dass es einen kurvilinearen Zusammenhang zwischen Geldbetrag und Rücklauf gibt, ist aus Sicht der Theorie subjektiver Werterwartung durchaus plausibel: Aus werterwartungstheoretischer Sicht erfolgt bei einem aus der subjektiven Perspektive der befragten Person hohen Geldbetrag ein Wechsel des kognitiven Rahmens für die soziale Definition der Befragungssituation (Esser 1996). Aus einer Situation des ‚good will‘ wird in der subjektiven Wahrnehmung der Zielperson ein Geschäftsakt, wenn der Sozialforscher den Geldbetrag zu hoch ansetzt. Diese sozialpsychologische Erklärung ist in unserem Fall kompatibel mit einer reaktanztheoretischen, der zufolge eine Befragungsperson versucht, ihre eingeengte Freiheit für die Wahl zwischen Teilnahme und Nichtteilnahme wieder herzustellen. So verweigern sie bewusst die Teilnahme, „um ihre durch zu hohe Incentives bedrohte persönliche Freiheit wieder herzustellen“ (Stadtmüller/Porst 2005: 5).

Antworten gegen eine Entschädigung) aufgefasst werden kann, wird es zu einer geringen Teilnahmebereitschaft führen (vgl. Fox/Crask/Kim 1988). Zudem wird sich die Befragungsperson unsicher darüber sein, ob sie den ausgelobten Geldbetrag auch tatsächlich erhält; schließlich ist ihr das Forscherteam in der Regel unbekannt. Wenn sie sich trotz der Bemühungen der Forscher (z. B. informatives Anschreiben) nicht sicher sein können, das Geld nach der Beantwortung der Fragen zu erhalten, dann werden sie in der Regel auch nicht kooperieren. Daher ist zu erwarten, dass Befragungspersonen, denen Geld versprochen wird, seltener teilnehmen als diejenigen, denen kein Geld oder Geld im Voraus geboten wird.⁷

(3) Des Weiteren wird bezüglich der Neigung für gesetzeswidrige Handlungen und der selbstberichteten Prävalenz – insbesondere des ‚ehrlichen Geständnisses‘ früherer Delinquenz – aus handlungs- und entscheidungstheoretischer Sicht argumentiert, dass es keinen zwingenden Grund gibt, weshalb Befragte, die monetäre Anreize im Voraus erhalten haben, bei Antworten vom ‚wahren Wert‘ abweichen sollten (vgl. Esser 1986; Reuband/Blasius 1996). Es wird daher angenommen, dass monetäre Anreize nur Auswirkungen auf die Ausschöpfung haben, aber keine auf die (wahrheitsgemäße) Beantwortung auch von Fragen zur Delinquenz. Naheliegender wäre – wie dies typischerweise bei Wahlbefragungen der Fall ist (Hartmeier/Fontana 2006) – ein so genanntes ‚Overreporting‘ von Straftaten bei den Personen, die bereits im Voraus mit einem Geldbetrag bedacht wurden. Aus Sicht des hier vertretenen aufgeklärten handlungs- und entscheidungstheoretischen Ansatzes ist dies jedoch recht unwahrscheinlich.

(4) Ebenso unwahrscheinlich ist aus dieser theoretischen Perspektive eine Stichprobenselektivität infolge eines im Voraus gewährten Geldbetrages. Entsprechend der Reziprozitätsthese sollte die Gabe für alle Bevölkerungsgruppen im gleichen Maße wirksam sein. Es ist daher zu erwarten, dass etwaige Unterschiede zwischen den Befragtengruppen und den Stichproben, die entweder monetäre Anreize erhielten oder denen diese versprochen wurden, zufällig sind (vgl. Reuband/Blasius 1996: 307).

7 Frey (1997) führt einen psychischen Mechanismus an, der dazu führen kann, dass monetäre ‚Incentives‘ zum gegenteiligen Ergebnis führen: „Wenn eine von außen kommende Intervention impliziert, dass die intrinsische Motivation der betreffenden Person nicht gewürdigt wird, bedeutet dies eine Missachtung ihrer Beweggründe. Die Person fühlt, dass ihr Engagement und ihre Kompetenz nicht geschätzt werden, was deren Wert in zwischenmenschlichen Interaktionen vermindert. Auf Grund der verminderten Selbsteinschätzung vermindert die betreffende Person ihren Einsatz“ (Frey 1997: 23). So könnte man annehmen, dass versprochene Gegenleistungen für einen beantworteten Fragebogen genau diese Reaktionen auslösen können, und somit ein ‚Response‘ eher wenig wahrscheinlich wird.

Formalisieren wir unsere Hypothesen, dann ergeben sich folgende Ungleichungen für die Wirkung einer Gabe im Voraus (Geldgeschenk) und einer versprochenen Gabe als Gegenleistung (Entschädigung) im Vergleich dazu, dass den Befragungspersonen nichts geschenkt oder versprochen wird (kein Treatment):

- (1) und (2) Ausschöpfung: Geldgeschenk > kein Treatment > Entschädigung
- (3) Antwortverzerrung: Geldgeschenk = kein Treatment = Entschädigung
- (4) Stichprobenselektivität: Geldgeschenk = kein Treatment = Entschädigung

3 Experimentelles Design

Das Methodenexperiment wurde im Rahmen des vom Schweizerischen Nationalfonds geförderten Projektes „*Determinanten des kriminellen Handelns: Die empirische Überprüfung eines erweiterten Rational Choice Modells anhand einer postalischen Befragung in Bern*“ und in Kooperation mit einem an der Technischen Universität Dresden angesiedelten und von der DFG geförderten Projekt mit der gleichen Thematik durchgeführt (Mehlkop/Becker 2007).⁸ Aus einer zufällig aus dem Einwohnermelderegister der Schweizer Bundeshauptstadt Bern gezogenen Datei (Einwohner Berns zwischen 18 und 65 Jahren) mit 4.000 Adressen wurden 510 zufällig gezogene Adressen nach dem Zufallsprinzip zu drei Gruppen mit jeweils 170 Personen aufgeteilt.

8 Ziel des Forschungsprojektes ist die empirische Überprüfung eines theoretischen Modells, das zu erklären versucht, warum und unter welchen Bedingungen Individuen bereit sind, Straftaten zu begehen. Nach der ökonomischen Theorie abweichenden Verhaltens ist Kriminalität als gesellschaftliches Phänomen das aggregierte Ergebnis absichtsvoller Entscheidungen und Handlungen von Individuen (Becker/Mehlkop 2006). Sie begehen dann Straftaten, wenn diese aus ihrer Sicht eher Vorteile (wie Wohlstand oder soziale Anerkennung) erbringen als legale Handlungen. In diesem Fall versuchen sie zielgerichtet über Straftaten, Gewinne zu erzielen, und achten dabei darauf, dass für sie selbst keine nachteiligen Konsequenzen (wie etwa Strafen, soziale Ächtung oder Selbstbeschädigung) entstehen. Dieses ökonomische Grundmodell vernachlässigt jedoch den Einfluss von individuellen Merkmalen wie etwa Fähigkeiten, Sozialisationseffekte, internalisierte Normen sowie subjektive Wahrnehmung und Evaluation von Gelegenheiten für Straftaten und ihre Erfolgswahrscheinlichkeiten, welche wiederum die Kosten-Nutzen-Abwägungen und damit den Entscheidungsprozess für oder gegen die Ausführung einer Straftat steuern (Becker/Mehlkop 2006). Diese Einflüsse variieren zwischen den Bildungsgruppen und sozialen Schichten. Aus theoretischer Sicht kann damit verdeutlicht werden, dass die häufig im Alltag vertretene Hypothese, weniger gebildete Personen oder untere und ärmere Sozialschichten seien krimineller als die ökonomisch privilegierten Sozialschichten, auf einem ökologischen Fehlschluss beruht. Berücksichtigt man die soziale Situation der Individuen, dann korrelieren Sozialmerkmale eher mit der Auswahl von ‚typischen‘ Straftaten und ihrer erfolgreichen Durchführung als mit der Häufigkeit von Straftaten. Mittels einer postalischen Befragung von über 3.000 Einwohner in Bern sollen Daten über deren delinquentes Handeln (Ladendiebstahl, Steuerhinterziehung, Schwarzfahren und Versicherungsbetrug) erhoben werden. Anhand dieser Daten soll das modifizierte ökonomische Modell der Kriminalität auf seine Tragfähigkeit überprüft und seine Erklärungskraft mit der von alternativen Erklärungsansätzen verglichen werden.

Der *ersten Vergleichsgruppe* (VG 1) wurde zu dem Fragebogen im Umfang von 20 Druckseiten (DIN A4 im maschinenlesbaren Layout sowie mit Foto des Bundeshauses und Bundesplatzes in Bern auf dem Deckblatt einschließlich einer kurzen Anleitung zum Ausfüllen des Fragebogens) und 88 Fragen eine 10-Franken-Banknote mit folgender Erläuterung beigefügt: „Als kleines Dankeschön für Ihre wertvolle Mitarbeit liegen diesem Schreiben 10 Franken bei. Wir freuen uns auf Ihre Teilnahme und danken Ihnen bereits an dieser Stelle bestens für Ihr Interesse und Ihre Zusammenarbeit“⁹. Der *zweiten Vergleichsgruppe* (VG 2) wurden 10 Franken als Dank für die Beantwortung des Fragebogens versprochen, die sie per Antwortkarte abrufen konnten: „Nach Erhalt der Postkarte senden wir Ihnen gerne ein kleines Dankeschön – nämlich 10 Franken – für Ihre wertvolle Mitarbeit zu. Wenn Sie die Postkarte zurückschicken, wissen wir zwar, dass Sie sich an unserer Umfrage beteiligt haben, nicht aber welcher der separat eingegangenen Fragebogen von Ihnen ausgefüllt wurde. Ihre Anonymität ist vollumfänglich gewährleistet. Ihre Adresse wird von uns gelöscht, sobald wir Ihnen die 10 Franken zugestellt haben“. Die *Kontrollgruppe* (KG) erhielt lediglich den Fragebogen inklusive Anschreiben, jedoch keine monetären Anreize.

Um diese drei Gruppen unterscheiden zu können, wurde jeder einzelne Fragebogen mit einer Nummer versehen. Diese dreistellige Zahl erschien auf jeder Seite des Fragebogens. Anhand dieser Nummer konnte der Fragebogen nach der Retournierung eindeutig einer der drei Untersuchungsgruppen zugeteilt werden. Neben der dreistelligen Nummer befand sich auf jeder Seite der Bogen auch noch eine weitere zehnstellige Zahl, welche für die von uns eingesetzte Scanner-Software (Teleform) notwendig war, die Fragebogen korrekt einzulesen. Die angeschriebenen Personen wurden im Begleitbrief darauf aufmerksam gemacht, dass die Fragebogen elektronisch erfasst werden und die auf dem Fragebogen aufgeführten Zahlen der korrekten Verarbeitung dienen. Sie wurden – entsprechend obigen Ausführungen – darüber aufgeklärt, wieso ihre Fragebogen – trotz angeführter Ziffern – anonym

9 Zwei Gründe sprechen für die Verwendung einer 10-Franken-Banknote. Zum einen ist dieser Geldschein der geringste Betrag in Papierform und damit leichter als eine Münze, wodurch Portogebühren eingespart werden können, die für die A-Post mit höchster Priorität preisintensiv sind. Zum anderen scheint der Betrag angemessen zu sein, während der nächste höhere Betrag in Papierform bereits 20 Franken beträgt und als Anreiz weniger effektiv ist. Diese Abstufung der Geldnoten hat uns davon abgehalten, im Experiment den Wert der ‚Incentives‘ zu variieren, um dadurch die Frage beantworten zu können, ob mit steigendem Wert der Gabe im Voraus die Ausschöpfungsquote zunimmt. In jüngster Zeit wurden bei einem ähnlichen Methodenexperiment in der Schweiz positive Erfahrungen mit dem Geldwert von 10 Franken gemacht (Diekmann/Jann 2001). Allerdings scheint heute bei einer weiten wie rasanten Verbreitung von Mobiltelefonen der Einsatz einer Telefonkarte für öffentliche Fernsprecher im Wert von 10 Schweizer Franken – wie dies Diekmann/Jann (2001) noch mit erheblichem Erfolg praktiziert haben – als materieller Anreiz wenig erfolgversprechend.

und nicht auf eine Person rückführbar sind. Lediglich eine einzige Befragungsperson entfernte diese Identifikationsziffern.

Die Feldzeit erstreckte sich in der Zeit vom 6. bis zum 21. März 2006. Da es uns nur um den unmittelbaren Effekt monetärer Anreize ging, wurde nicht nachgefasst. Der Fragebogen wurde den Befragten per A-Post, dem schnellsten Postweg in Bern mit einer Zustelldauer von einem Tag, zugestellt. Nach Ausfüllen des Fragebogens konnten die Befragten diesen im beigelegten adressierten und frankierten Briefkuvert zurücksenden.¹⁰

4 Empirische Resultate

4.1 Zusammenhang von monetären Anreizen und der Ausschöpfung

In der Tat gibt es einen signifikanten Zusammenhang zwischen der Strategie, den ‚Response‘ mittels monetärer Anreize zu erhöhen, und dem Rücklauf von ausgefüllten Fragebögen (Tabelle 1). Damit werden die zuvor diskutierten Befunde identischer oder ähnlicher Methodenexperimente repliziert (vgl. Diekmann/Jann 2001; Mehlkop/Becker 2007).

Tabelle 1 Ausschöpfungsquote einer postalischen Befragung in der Stadt Bern zum Thema Kriminalität in Abhängigkeit von monetärem Anreiz (Zeitraum: 16 Tage)

	Netto-Stichprobe	Neutrale Ausfälle	Brutto-Stichprobe	Keine Teilnahme	Teilnahme	Rücklauf
Kontrollgruppe	170	3	167	102	65	38,9%
VG 1: Beigelegt	170	5	165	61	104	63,0%
VG 2: Versprochen	170	5	165	109	56	33,9%
Insgesamt	510	13	497	272	225	45,3%

$\Lambda = 0,181$

10 Um eine vordergründige Sensibilisierung der Befragten auf das Untersuchungsthema, nämlich Determinanten der individuellen Delinquenz zu vermeiden, wurde die Studie unter das Motto „Lebensqualität und Kriminalität in Bern“ gestellt. Dieses Motto greift auch derzeitige Diskussionen in der Berner Öffentlichkeit auf. Für den Einstieg wurden Sachverhalte wie dringende Probleme der Stadt Bern, Einschätzung der Arbeit der Berner Polizei und der öffentlichen Sicherheit, eigene Viktimisierung sowie Fragen zum Lebensalltag und zu sozialen Normen herangezogen (vgl. Porst 1999).

Anhand der Versuchsbedingung kann der Vorhersagefehler für die Ausschöpfungsquote um 18% reduziert werden. Sie beträgt – ohne jegliches Nachfassen und abzüglich der Ausfälle – insgesamt rund 45% und ist für die Gruppe der Befragten, die das Geld im Voraus erhielten, mit 63% am höchsten, gefolgt von der Kontrollgruppe (39%).¹¹ Am geringsten ist mit einer Quote von 34% – wie ebenfalls theoretisch erwartet – die Ausschöpfung für die Gruppe der Befragten, denen das Geld versprochen wurde (vgl. Diekmann/Jann 2001).¹² Von den 56 Personen aus der VG 2 haben 49 Personen die Karte rechtzeitig retourniert und als Gegengabe auch die 10 Franken erhalten. Da weniger Personen die Karte als den Fragebogen zurück geschickt haben, ist die Wahrscheinlichkeit, dass sich Personen Geld erschlichen haben, gering.

Die (mittels logistischer Regression geschätzte) relative Chance, einen ausgefüllten Fragebogen zu erhalten, ist mit Geld als bedingungslose Vorleistung mehr als doppelt so hoch als ohne jeglichen Anreiz (Tabelle 2). Bei Kontrolle der Zeitabhängigkeit des Rücklaufs ist die Teilnahmebereitschaft der Personen, die bereits im Voraus den Geldbetrag von 10 Franken erhalten haben, um $(2,039 - 1) * 100\% = 104\%$ höher pro Zeiteinheit als die der Kontrollgruppe. Die Rücklaufquoten sind bei den Befragten, denen das Geld im Tausch gegen den ausgefüllten Fragebogen versprochen wurde, niedriger als für die Kontrollgruppe, aber die Unterschiede sind statistisch nicht signifikant. Ohne dass wir die tatsächlichen sozialen und psychischen Mechanismen kennen, die dem Rücklauf unter verschiedenen Versuchsbedingungen zugrunde liegen, kann jetzt schon festgehalten werden, dass – gemessen am Ergebnis – die austausch- und werterwartungstheoretischen Hypothesen durch unsere Befunde gestützt wer-

11 Im Vergleich dazu seien die Rücklaufquoten für ein ähnliches Methodenexperiment genannt, das wir in Dresden durchgeführt haben (Mehlkop/Becker 2007). Insgesamt lag in Dresden der Rücklauf bei rund 40%. Rund 52% der im Voraus mit Geld ‚beschenkten‘ Befragungspersonen schickten den Fragebogen zurück, aber nur 28% der Befragten in der Kontrollgruppe. Eine Vergleichsgruppe, der das Geld gegen Leistung versprochen wurde, gab es nicht. Es sei nochmals betont, dass diese Ausschöpfung bei mit einem heiklen Thema, einem relativ langen Fragebogen und ohne Nachfassen erreicht wurden. Für die Ausfälle ist noch zu erwähnen, dass fünf der 170 Berner Befragungspersonen in der ersten Experimentalgruppe das Geld zusammen mit dem unausgefüllten Fragebogen zurückgeschickt haben. Bei den anderen acht Ausfällen war die Adresse nicht mehr aktuell oder der Adressat verstorben. Ähnliches kann für das Experiment in Dresden berichtet werden: Fünf von 100 Personen aus der Gruppe mit Geldgeschenk haben den Fragebogen zwar unausgefüllt, aber mit der 5-Euro-Banknote zurückgeschickt, und eine Person hat einen ausgefüllten Fragebogen zusammen mit der 5-Euro-Banknote zurückgeschickt (Mehlkop/Becker 2007).

12 Dass unsere Rücklaufquoten etwas niedriger sind als beim Experiment von Diekmann/Jann (2001), hat mehrere Gründe. Nicht nur, dass wir mit einer sensiblen und heiklen Thematik an die Befragungspersonen herangetreten sind, sondern es ist zu berücksichtigen, dass Diekmann/Jann (2001) auf eine selektive Stichprobe zurückgreifen. Es handelt sich bei den Autoren um eine Nachbefragung von Personen, die bereits telefonisch interviewt wurden. In unserem Fall hingegen liegt eine Erstbefragung vor. Auch ist unsere Feldzeit mit 16 Tagen weitaus kürzer.

den. Die Hypothese der strikten Rationalität hält wiederum – wie in anderen Studien auch – nicht der Empirie stand.

Tabelle 2 Gruppenunterschiede in der Ausschöpfung in Abhängigkeit von monetärem Anreiz (odds ratios)

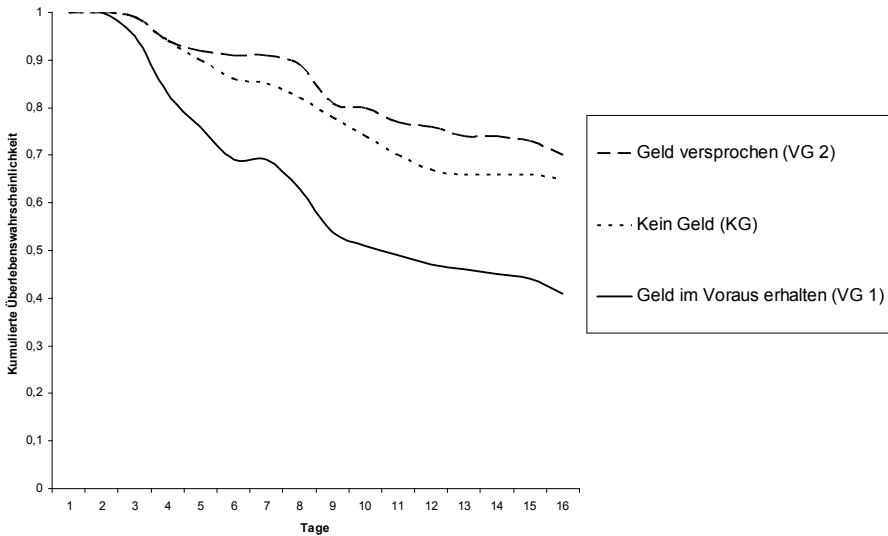
	Logistische Regression	Cox-Regression
Kontrollgruppe	1	1
VG 1: Beigelegt	2,675***	2,039***
VG 2: Versprochen	0,806	0,821
Pseudo-R ² (McFadden)	0,047	0,013
N	497	497

*** $p \leq 0,001$

Unseren Befunden zufolge dürfte der Einsatz von monetären Anreizen nicht nur effektiv, sondern auch kostengünstiger sein als andere Strategien und Maßnahmen, um den Rücklauf zu erhöhen. Denn je mehr Personen im Zeitfenster vor dem Nachfassen den Fragebogen beantworten und zurücksenden, desto kleiner wird die Risikomenge derjenigen, die nochmals angeschrieben werden müssen und daher auch Kosten verursachen. Diese Vermutung wird zunächst anhand der mittleren Dauer des Rücklaufs überprüft, da davon der Einsatz von Nachfassaktionen im Sinne der TDM nach Dillman (2000) abhängt.

Die in Abbildung 1 dargestellten Überlebenswahrscheinlichkeiten – die relativen Anteile derjenigen, die zu den einzelnen Zeitpunkten den Fragebogen noch nicht retourniert haben (Blossfeld/Rohwer 1995) – unterstützen unsere vorherigen Befunde. So erfolgt im Beobachtungszeitraum von 16 Tagen der Rücklauf am ehesten für die Gruppe, die Geld im Voraus erhalten hat, gefolgt von der Kontrollgruppe und dann erst der Gruppe, der Geld nach erbrachter Leistung in Aussicht gestellt wurde. Bis ein Viertel der Fragebogen zurückkommt, und damit noch drei Viertel der ‚Risikomenge‘ überlebt haben, vergehen bei der im Voraus beschenkten Gruppe 5 Tage, bei der Gruppe, der das Geld versprochen wurde, knapp 11 Tage und schließlich bei der Kontrollgruppe 9 Tage. Die Zeitdifferenzen sind entsprechenden Tests – wie etwa Log-Rank-Test oder Wilcoxon-Test – zufolge signifikant. Nur für die Befragungspersonen, denen zum Fragebogen eine 10-Franken-Banknote beigelegt wurde, kann ein Median von 9 Tagen errechnet werden. Am wenigsten effizient und effektiv erweist es sich dabei wiederum, den Befragungspersonen Geld zu versprechen.

Abbildung 1 Überlebenswahrscheinlichkeiten für den Rücklauf in Abhängigkeit von monetären Anreizen



Dieser Befund kann mit einem allgemeinen handlungstheoretischen Ansatz wie der ‚Frame Selection Theory“ (Esser 1996; Kroneberg 2007) differenziert beschrieben werden: Personen, welche die Reziprozitätsnorm internalisiert haben, handeln nach dieser, indem sie bei gegebener Vorleistung einen kognitiven Rahmen aktivieren, der die Gegenleistung als adäquate Handlungsweise („habit“) beinhaltet (vgl. Esser 1990): Folglich wird der Fragebogen sofort nach Erhalt und ohne viel Nachdenken wahrheitsgemäß ausgefüllt und rasch zurückgesendet. Bei Personen, denen das Geld lediglich versprochen wurde, sollte der kognitive Reziprozitätsrahmen nicht greifen. Das sollte deswegen nicht der Fall sein, weil das in Aussicht gestellte Geld an sich keine Reziprozität beinhaltet, auf die reagiert werden muss. Daher wird der kognitive ‚Frame‘ für Reziprozität nicht aktiviert und die Reziprozitätsnorm kommt nicht zum Tragen. Das versprochene Geld ist – wie bereits im theoretischen Teil ausgeführt – das Angebot für einen Tausch ‚Geld gegen Antwort‘. Der Tausch kann von der Befragungsperson akzeptiert oder abgelehnt werden, aber an sich generiert der Tausch keine Reziprozität. Da dann der angemessene ‚reciprocity frame‘ nicht aktiviert wird, sollte in der VG 2 (Geld versprochen) ein mehr oder weniger lang andauerndes Rasonieren über Kosten und Nutzen einer Teilnahme eine deutlich größere Rolle spielen als in der VG 1 (Geld im Voraus). Wegen eines fehlenden kognitiven Rahmens dauert die Evaluation von Handlungsalternativen („Teilnahme“

oder ‚Verweigerung‘) und Entscheidung für eine bestimmte Handlung vergleichsweise länger. Anhand der Überlebenswahrscheinlichkeiten in Abbildung 1 sehen wir entsprechende indirekte Hinweise für diese Ad hoc-Interpretation: Da die Personen in der VG 2 (Geld versprochen) bei einem fehlenden kognitiven Rahmen lange über eine angemessene Reaktion auf den postalischen Fragebogen nachdenken müssen, brauchen sie vergleichsweise länger, um den Fragebogen zurückzuschicken. Wohl gemerkt können wir keinen direkten Test für diese Interpretation liefern, da wir die entscheidenden Mechanismen wie kognitive Rahmen und gespeicherte Handlungs-routinen nicht erhoben haben. Zudem müssten wir diese auch bei den Nonrespondenten erheben.

Ungeachtet dessen werden insgesamt die Befunde von Brennan (1992), Harkness et al. (1998) sowie von Arzheimer/Klein (1998) unterstützt. Unsere Befunde für die Dauer, bis eine entsprechende Rücklaufquote erreicht wird, sprechen dafür, dass über Geld im Voraus erhebliche organisatorische Aufwendungen und Kosten eingespart werden können, die wegen Nachfassaktionen auch bei denen entstehen würden, die zur Antwort bereit sind, sich aber mit dem Rücksenden Zeit lassen (Berger 2006: 92).¹³

4.2 Zusammenhang von monetären Anreizen und Datenqualität

Auch hinsichtlich der Neigung wie Prävalenz für Delinquenz werden unsere Annahmen empirisch unterstützt (Tabelle 3 und 4). Fehlende Werte zu möglichen Straftaten wie Ladendiebstahl, Steuerhinterziehung, Versicherungsbetrug oder Schwarzfahren sind zufällig über die Kontroll- und Vergleichsgruppen verteilt. Daher werden sie nicht weiter berücksichtigt. Zwar neigen einerseits Personen, denen entweder ein Geldbetrag im Voraus zugeschickt oder versprochen wurde, eher als die Kontrollgruppe dazu, keinen Ladendiebstahl, keinen Steuerbetrug oder keine Schwarzfahrten zu begehen. Andererseits würden sie eher als Personen in der Kon-

13 Unter dem Kostengesichtspunkt – und wir sehen, dass mit monetären Vorleistungen der Rücklauf rascher erfolgt und hierbei eine Kostenersparnis gegeben ist, wenn weitere Schritte der TDM nicht mehr notwendig sind – wäre schließlich noch der Schwellenwert für die ‚Incentives‘ auszuloten. Bei Geldbeträgen von 5 Euro oder 10 Franken pro Befragungsperson müssen in Forschungsanträgen gewisse Überzeugungsarbeiten geleistet werden, da bei größeren Stichproben höhere Summen als ‚Geschenke‘ eingesetzt werden. In zukünftigen Methodenexperimenten könnten die Geldbeträge abgestuft zwischen 0 bis 100 Euro bzw. Franken variiert werden und in Beziehung zum Rücklauf gesetzt werden. Dadurch lässt sich beurteilen, ob es einen Grenznutzen für die Befragten gibt, und ob Forscher einen Schwellenwert (die zweite Ableitung des geschätzten Grenznutzens) berücksichtigen müssen. Vom Nutzen solcher Methodenexperimente sind dann die Sponsoren zu überzeugen.

trollgruppe ihre Versicherung betrügen. Aber diese Unterschiede sind statistisch insignifikant und daher zufällig.

Da es sich den drei miteinander verglichenen Gruppen um Zufallsstichproben handelt, sind Abweichungen vom ‚wahren Wert‘ bei allen Personen gleich wahrscheinlich. Bereits gewährte oder versprochene Geldbeträge sind keine kausalen Faktoren, die wegen sozialer Erwünschtheit oder ‚sponsorship effects‘ zum systematischen ‚Overreporting‘ von intendierten gesetzeswidrigen Handlungen führen.

Tabelle 3 Neigung zu intendierten Straftaten
(odds ratios, geschätzt mit logistischer Regression)

	Ladendiebstahl	Steuerbetrug	Versicherungsbetrug	Schwarzfahren
Kontrollgruppe	1	1	1	1
VG 1: Beigelegt	0,6	0,6	1,2	0,9
VG 2: Versprochen	0,6	0,5	1,7	0,6
Pseudo-R ² (McFadden)	0,009	0,013	0,007	0,010
N	199	200	199	197

In ebenfalls unsystematischer Weise variieren die Berichte der Befragten über von ihnen bereits begangene Straftaten (siehe Modell 1 in Tabelle 4). Zwar geben diejenigen Personen, die bereits den Geldbetrag erhalten haben, in einem geringeren Maße an, bereits Gesetze übertreten zu haben, als die anderen beiden Gruppen. Jedoch sind – auch bei Kontrolle des Alters (siehe Modell 2) – die Unterschiede der Prävalenzraten zwischen den drei Gruppen wiederum statistisch insignifikant. Weiterführende Analysen haben ergeben, dass sich die Kontroll- und Vergleichsgruppen auch nicht hinsichtlich von Merkmalen (Gesetzestreue, Achtung von Eigentumsrechten, materialistische Orientierung, etc.) unterscheiden, die mit Delinquenz korrelieren (Becker/Imhof/Raimann 2007).

Erste Analysen der Hauptuntersuchungen für Bern und Dresden belegen, dass Neigung wie Prävalenz für Straftaten mit der Akzeptanz sozialer Normen (wie etwa Eigentumsrechte, Reziprozität etc.) korrelieren (Becker/Imhof/Raimann 2007): Personen, die solche Normen internalisiert haben, neigen bei Kontrolle subjektiv erwarteter Nutzen und Kosten sowie subjektiv erwarteter Erfolgs- und Entdeckungswahrscheinlichkeiten signifikant seltener zu delinquentem Handeln.

Tabelle 4 Prävalenz für Straftaten
(odds ratios, geschätzt mit logistischer Regression)

Modell	Ladendieb- stahl		Steuerbetrug		Versicherungs- betrug		Schwarzfahren	
	1	2	1	2	1	2	1	2
Kontrollgruppe	1	1	1	1	1	1	1	1
VG 1: Beigelegt	0,6	0,5	0,3	0,3	0,3	0,2	0,9	0,9
VG 2: Versprochen	1,1	1,2	0,1	0,1	1,2	1,1	0,5	0,5
Alter		1,2*		1,2		1,5		1,0
Alter ²		0,9*		1,1		0,9		1,0
Pseudo-R ² (McFadden)	0,016	0,067	0,076	0,097	0,046	0,094	0,013	0,126
N	199	196	199	196	199	196	199	196

* $p \leq 0,05$

Im Sinne der ‚Frame Selection Theory‘ nach Esser (2003) und des darauf basierenden Modells wertrationalen Handelns (Kroneberg 2007) deliberieren diese Personen nicht über eigenes delinquentes Verhalten, sondern streichen illegale Handlungsalternativen aus ihrem Set wählbarer Handlungen und prozessieren quasi-automatisch ausschließlich legale Handlungsalternativen. Demzufolge wäre umgekehrt nicht ausgeschlossen, dass delinquente Personen, die durch Gewährung eines ‚prepaid monetary incentives‘ mit der Reziprozitätsnorm konfrontiert werden, eher dazu neigen, das Geld einzustecken ohne mit der Beantwortung des Fragebogens eine Gegenleistung zu liefern. Personen, die eine Reziprozitätsnorm internalisiert haben, reagieren auf die ‚Incentives‘ mit einem kognitiven ‚Frame‘: Das Geld wird als Vorleistung Anderer interpretiert und damit der Bezugsrahmen ‚Gegenleistung‘ aktiviert. Der Fragebogen wird ausgefüllt und zurückgesendet. Menschen, die hingegen die Reziprozitätsnorm nicht internalisiert haben, wägen Kosten und Nutzen der Teilnahme bzw. den Wert von monetären Gaben ab. Die Vermutung, ob insbesondere Personen mit delinquentem Verhalten die Gabe im Voraus zwar kassieren, aber nicht antworten, kann allerdings nur mit Informationen über Nonrespondenten überprüft werden. Diese liegen in unserem Fall nicht vor. Sollte diese Vermutung zutreffen, dann wäre bei einem Einsatz von ‚prepaid monetary incentives‘ nicht ausgeschlossen, dass sowohl die realisierte Stichprobe selektiv ist als auch die Neigung wie Prävalenz für Straftaten systematisch unterschätzt wird. Aber unsere Befunde belegen zumindest für die Respondenten eindeutig: Sowohl bei der Neigung als auch für die Prävalenz führen monetäre Anreize bei den Befragten *nicht* zu verzerrten Angaben über eigene Delinquenz.

4.3 Zusammenhang von monetären Anreizen und Stichprobenzusammensetzungen

Auch wenn monetäre Anreize nicht mit systematischem Antwortverhalten einhergehen, ist nicht ausgeschlossen, dass sie aufgrund der signifikanten Unterschiede in den Ausschöpfungen zu selektiven Stichproben führen. Da wir in der Haupterhebung dem Fragebogen einen Geldbetrag von 10 Franken im Voraus beifügen, ist es interessant, zu wissen, ob sich die Gruppe der im Voraus Beschenkten in Bezug auf sozialstrukturelle Merkmale und Korrelate mit Delinquenz von den anderen beiden Gruppen unterscheidet. Sollte dies der Fall sein, könnte dies auch ein Problem für die Interpretation der erhobenen Daten bedeuteten.

Allerdings ist eine Beurteilung mit gewissen Unsicherheiten behaftet. Erstens ist die Verteilung von Einstellungen und Werthaltungen zur Delinquenz in der Grundgesamtheit unbekannt, so dass es nicht möglich ist, Abweichungen in den Stichproben beurteilen zu können. Das ist jedoch für unsere Problemstellung unerheblich. Zweitens ist die Verteilung dieser Strukturen auch für die Ausgangsstichproben – sprich für die Vergleichs- und Kontrollgruppen – unbekannt. Aus stichprobentheoretischen Gründen können wir nur vermuten, dass aufgrund der zufälligen Ziehung dieser Teilstichproben keine systematischen Unterschiede zwischen diesen bestehen. Drittens haben wir die Schwierigkeit, dass wir unterstellen müssen, dass etwa die Referenzgruppe – in unserem Fall ist es die Gruppe, die kein ‚treatment‘ erhielt – diejenige ist, bei der am ehesten eine Stichprobe realisiert werden kann, die im oftmals missverstandenen Methodenjargon als ‚repräsentativ‘ bezeichnet wird (vgl. Pötter/Rendtel 1993). Daher haben folgende Analysen, die *nicht* mit so genannten Repräsentativitätsstudien zu verwechseln sind, eher einen explorativen Charakter.

Zunächst ist festzustellen, dass in Bezug auf sozialstrukturelle Merkmale wie etwa Alter, Geschlecht, Nationalität, Bildungsniveau und Schichtzugehörigkeit keine signifikanten Unterschiede zwischen den Gruppen, denen entweder Geld beigelegt oder versprochen wurde, und den Befragten, die keine monetäre Anreize erhielten, bestehen (Modell 1 in Tabelle 5).¹⁴ Gleiches gilt für Bewertungen von Werthaltungen und Verhaltensweisen wie etwa Beitrag zum Allgemeinwohl, Achtung der Eigentumsrechte anderer Personen, Verwerflichkeit von Bereicherung auf Kosten anderer oder strikte Gesetzestreue (Modell 2 in Tabelle 5).

14 Allerdings besagt dieser Befund nicht, inwieweit die Sozialstruktur der realisierten Teilstichproben mit der in der Grundgesamtheit übereinstimmt. So wissen wir beispielsweise nicht, ob die realisierten Stichproben durch einen Bildungsbias gekennzeichnet sind (Hartmann/Schimpl-Neimanns 1993). Mit einiger Sicherheit kann jedoch gemutmaßt werden, dass alle drei Teilstichproben in ihrer sozialstrukturellen Zusammensetzung in gleichem Maße mehr oder weniger gut mit der Grundgesamtheit übereinstimmen.

Andere Befunde liegen für die Bewertung von ausgewählten Strafdelikten wie etwa Ladendiebstahl, Versicherungsbetrug, Schwarzfahren oder Steuerhinterziehung vor (Modell 3 in Tabelle 5). Einerseits sehen Personen, denen Geld als Gegenleistung in Aussicht gestellt wurde, einen Versicherungsbetrug als weniger schlimm an als Personen, die keine monetären Anreize erhielten. Andererseits finden Personen, denen ein Geldbetrag versprochen wurde, das Schwarzfahren verwerflicher als die Kontrollgruppe. Die Personen, die einen Geldbetrag im Voraus erhielten, unterscheiden sich bei ihren Bewertungen nicht signifikant von der Kontrollgruppe.

Tabelle 5 Selektivität der Stichproben durch monetäre Anreize
(odds ratios, geschätzt mit multinomialer Logit-Regression)

Modell	GB	GV	GB	GV	GB	GV
	1		2		3	
<i>Sozialstruktur</i>						
Alter	1,0	1,0				
Mann	1	1				
Frau	1,2	1,1				
Nicht-Schweizer	1	1				
Schweizer	0,8	0,9				
Niedrige Bildung	1	1				
Mittlere Bildung	1,3	0,3				
Hohe Bildung	1,6	0,7				
Arbeiterschichten	1	1				
Mittelschichten	0,7	1,9				
Höhere Schichten	0,5	0,5				
<i>Einstellung zu...</i>						
Beitrag zum Allgemeinwohl			0,9	1,2		
Schutz der Eigentumsrechte			0,7	0,7		
Egoistische Bereicherung			1,5	1,0		
Gesetzestreue			0,9	1,3		
<i>Negative Bewertung von...</i>						
Ladendiebstahl					0,8	1,2
Versicherungsbetrug					1,1	0,6*
Schwarzfahren					1,1	1,5*
Steuerhinterziehung					0,9	1,3
Pseudo-R ² (McFadden)	0,041		0,025		0,026	
N	194		200		199	

GB = Geld beigelegt (d. h. Gabe im Voraus) und GV = Geld versprochen (d. h. Entschädigung in Aussicht gestellt)

* $p \leq 0,05$

Folgendes Fazit lässt sich aus diesen Befunden ziehen: Wenn davon ausgegangen wird, dass ein monetärer Anreiz eher zu selektiven Stichproben führt als ein ‚non-treatment‘, dann weisen unsere Befunde darauf hin, dass dies nur in dem Falle zutreffend ist, wenn ein monetärer Anreiz als *Entschädigung für das Ausfüllen und Rücksenden des Fragebogens in Aussicht gestellt wird*. Hingegen resultiert ein bedingungslos im Voraus beigelegter Geldbeitrag *nicht* – wie man dies eher erwarten würde – in selektiven Stichproben. Diese Befunde gelten auch für eine Vielzahl anderer Items (Deprivation, Anomie oder psychische Dispositionen wie etwa Spontaneität, Risikobereitschaft, Empathie, Egozentrik, Sublimation etc.), die wir aus Gründen der Redundanz nicht explizit präsentieren.

5 Zusammenfassung und Diskussion

Ziel des vorliegenden, im Rahmen einer Bevölkerungsumfrage in Bern durchgeführten Methodenexperimentes war es, für eine postalische Befragung empirisch festzustellen, ob materielle Anreize auch bei sensiblen und für die Befragungspersonen möglicherweise heiklen Fragen zur eigenen Delinquenz (Ladendiebstahl, Steuerhinterziehung, Schwarzfahren und Versicherungsbetrug) die Rücklaufquote in bedeutsamer Weise fördern. In der experimentellen Versuchsanordnung wurden drei Gruppen unterschieden: Der ersten Untersuchungsgruppe wurde dem Fragebogen eine 10-Franken-Banknote beigelegt; der zweiten Untersuchungsgruppe hingegen wurden 10 Franken versprochen, wenn sie den Fragebogen beantworten und termingerecht zurückschicken. Die Kontrollgruppe erhielt kein zusätzliches ‚treatment‘. Aus der Perspektive der strikten, d. h. ökonomistischen Rationalität war theoretisch zu erwarten, dass die monetäre Vorleistung nicht zur Steigerung des Rücklaufs beiträgt, da die Befragungspersonen das Geld nehmen, aber keine Gegenleistung erbringen. Hingegen wäre es für die Befragten, denen Geld versprochen wurde, unter der Voraussetzung, die 10 Franken kompensieren ihre Mühen und Anstrengungen, auch rational, den Fragebogen auszufüllen, um an das Geld zu gelangen. In diesem Falle hätten versprochene monetäre Anreize eine positive Wirksamkeit auf den Rücklauf bei postalischen Befragungen. Jedoch belegen empirische Studien das Gegenteil: In der Regel beantworten Personen, die bedingungslos im Voraus Geld erhalten haben, den Fragebogen eher als Personen, denen Geld für den Fall der Antwort versprochen wurde. Oftmals ist der Rücklauf bei der letzten Gruppe sogar niedriger als für eine Kontrollgruppe, denen kein monetärer Anreiz zukam.

Dieses scheinbare Paradox der Teilnahme an einer wissenschaftlichen Befragung löst sich auf, wenn austausch- und werterwartungstheoretische Erklä-

rungsansätze herangezogen werden. Aus dieser theoretischen Perspektive wird angenommen, damit mit einer monetären Vorleistung seitens der Forscher eine Verpflichtung zur Reziprozität bei den Befragungspersonen entsteht, deren Einhaltung einen Nutzen für sie als kooperative Akteure erbringt. Eine bedingungslose monetäre Vorleistung stellt offensichtlich eine vertrauensbildende Maßnahme dar, die notwendig scheint, damit eine befragte Person ihr unbekanntem Forschern (sensible) Informationen auf gezielte Frage zur Verfügung stellt. Versprochenes Geld ist dagegen kontraproduktiv: Da Befragungspersonen sich trotz des Versprechens der Forscher nicht sicher sein können, das Geld nach Beantwortung der Fragen tatsächlich zu erhalten, werden sie in der Regel auch nicht kooperieren.

Auch wenn wir die theoretisch angeführten sozialen und psychischen Mechanismen nicht direkt gemessen haben, und das ist in jedem der uns bekannten Methodenexperimente ebenfalls das Problem, so werden – gemessen an den ‚treatments‘ und ‚outcomes‘ – unsere Modellüberlegungen durch die vorgelegten empirischen Befunde gestützt. Sofortige monetäre Leistungen fördern auch bei Fragen zur Delinquenz den Rücklauf einer postalischen Befragung. Ohne Nachfassen antworteten innerhalb von zwei Wochen bis zu zwei Drittel der Befragten, denen 10 Schweizer Franken zum Fragebogen beigelegt wurden. In Kombination mit zwei Nachfassaktionen gemäß der ‚Tailored Design Method‘ (TDM) von Dillman (2000) wäre eine zufriedenstellend hohe und von geringerer Selektivität gekennzeichnete Rücklaufquote zu erwarten. Möglicherweise ist diese Einschätzung sehr optimistisch. Denn mit zunehmender Dauer, mit der eine Vorleistung zurückliegt, verblasst möglicherweise das ‚Gefühl der Reziprozität‘.¹⁵ In einigen Fällen dürfte eine Auffrischung über ein Erinnerungsschreiben kaum gelingen, so dass die Rücklaufquote unter dem erwartbaren Optimum liegen wird.

Offensichtlich gibt es auch keine unbeabsichtigten Einflüsse beim Antwortverhalten der Befragten. Ein systematisches ‚overreporting‘ von bereits begangenen

15 Dass diese Einschätzung nicht zu optimistisch ist, zeigen die Rückläufe für unsere in der Zwischenzeit abgeschlossene Haupterhebung. Von 2.854 Fragebögen der Brutto-Stichprobe (neutrale Ausfälle: 5,2%) kamen nach zweimaligem Nachfassen insgesamt 2.230 verwertbare Fragebögen in der Feldzeit zurück. Die Ausschöpfungsquote beträgt somit rund 78%. Davon haben 41 Personen (fast 2%) zusätzlich zum ausgefüllten Fragebogen auch die 10 Franken zurückgeschickt. Von den 624 Verweigerern schickte mehr als die Hälfte (343 Personen) den Geldbetrag von 10 Franken zusammen mit dem leeren Fragebogen zurück (unter anderem mit Bemerkungen, dass sie unentgeltlich mitmachen wollen oder dass wir Forscher uns davon einen Kaffee leisten sollten). Nach der Feldzeit erreichten uns 70 teils ausgefüllte, teils leere Fragebögen. Auch hierbei sandten 17 Befragungspersonen den Geldbetrag zurück; eine davon – aus welchen Gründen auch immer – sogar den doppelten Betrag von 20 Franken. Die Retournierung der Geldbeträge nach nicht erbrachter Leistung spricht für die Reziprozitätsthese an sich und auch für das austauschtheoretische Argument, dass sich Befragte in einem unerwünschten Abhängigkeitsverhältnis sehen, wenn sie nicht kooperieren.

Straftaten oder der Prävalenz infolge sozialer Erwünschtheit, die wiederum durch monetäre Anreize ausgelöst werden könnte, konnte nicht festgestellt werden. Verweigerungen von Fragen zu Straftaten hängen auch nicht davon ab, ob monetäre Anreize (nicht) gewährt oder versprochen wurden. Ebenso gibt es keine Hinweise dafür, dass bedingungslos im Voraus gewährte Anreize zu selektiven Analysestichproben führen.

Unsere Befunde unterstützen bereits vorliegende Methodenexperimente (Mehlkop/Becker 2007; Berger 2006) und belegen darüber hinaus, dass es auch für eine Studie, in der die Befragungspersonen aufgefordert werden, selbst über ihr rechtswidriges Verhalten zu berichten, sinnvoll wäre, monetäre Anreize sofort und bedingungslos zu liefern. Allerdings sind wir davon überzeugt, dass der überraschend deutliche Erfolg unserer Bemühungen, eine hohe Rücklaufquote zu erzielen, auch auf sorgfältigen Vorarbeiten basiert. Der zuletzt eingesetzte Fragebogen wurde zwei kognitiven Pretests unterzogen. Des Weiteren wurde er in einer webbasierten Online-Befragung unter Berner Studierenden getestet (Becker/Imhof/Raimann 2007). Sowohl bei der Gestaltung des Anschreibens als auch des Fragebogens wurden die Empfehlungen von Stadtmüller und Porst (2005), Porst (2000, 2001) und von Dillman (2000) umgesetzt. Offensichtlich war die zugesicherte Anonymität überzeugend und auch die Erläuterung der Kennziffern auf dem oberen Rand des Fragebogens als Justierungs- und Erkennungspunkte für das maschinelle Einlesen erschien den Befragungspersonen transparent. Und schließlich wurde an die Kooperationsbereitschaft der Befragungsperson appelliert und die Bedeutsamkeit ihres Beitrags für die Sozialwissenschaft im Allgemeinen und für die schweizerische Forschung im Besonderen hervorgehoben.

In theoretischer Hinsicht sind die zentralen Mechanismen weiterhin ungeklärt, *warum* Befragungspersonen eher antworten, wenn sie beim Erstkontakt zusätzlich zum Fragebogen bedingungslos im Voraus Geld erhalten, als wenn ihnen ein Geldbetrag für den Fall in Aussicht gestellt wird, dass sie den Fragebogen ausfüllen und zurückschicken. Zwar sprechen die vorliegenden Befunde für die Theorie subjektiver Werterwartung im Allgemeinen und für die Reziprozitätsthese im Besonderen. Aber das ist noch lange keine Gewähr dafür, dass diese Theorien zutreffend sind, wenn nicht empirisch nachgewiesen wird, dass die prognostizierten Mechanismen sowohl der Kosten-Nutzen-Abwägung als auch der Reziprozität tatsächlich für den Rücklauf verantwortlich sind (Hedström/Swedberg 1998). Deren genaue Kenntnis – also der Mechanismen und Prozesse in der ‚black box‘ – bedient nicht einen Methodenfetischismus oder eine theoretische Eitelkeit nach dem Motto: „Haben wir doch schon immer gewusst“. Vielmehr kann dieses Wissen dazu beitragen, den Rücklauf im Weberschen Sinne ursächlich zu verstehen und

dadurch auch besser kontrollieren und steuern zu können. Damit wäre die Möglichkeit eingeschlossen, die Ausschöpfung bei gleichzeitiger Vermeidung von Selektivitäten noch weiter zu steigern als dies bislang der Fall ist. Dazu reicht es sicherlich nicht aus, lediglich sozialstrukturelle Merkmale der Verweigerer in Erfahrung zu bringen. Vielmehr benötigen wir für alle Befragungspersonen exakte Informationen über ihre Gründe für Kooperation und Verweigerung. Erst dann können weitere Maßnahmen ergriffen werden, um auch potentielle Verweigerer zur Teilnahme an postalischen Befragungen zu motivieren oder gar Verweigerer zu Teilnehmern zu konvertieren (Neller 2005). Als erwünschter Nebeneffekt könnte sich eine geringere Selektivität von realisierten Stichproben ergeben. Darauf kommt es eher an als bloß auf die quantitative Ausschöpfung (Koch 1998; Pötter/Rendtel 1993).

Literatur

- Armstrong, J.S., 1975: Monetary Incentives in Mail Surveys. *Public Opinion Quarterly* 39: 111-116.
- Arzheimer, K. und M. Klein, 1998: Die Wirkung materieller Incentives auf den Rücklauf einer schriftlichen Befragung. *ZA-Information* 43: 6-31.
- Becker, R. und G. Mehlkop, 2006: Social Class and Delinquency. *Rationality & Society* 17: 195-235.
- Becker, R., 2006: Selective Response to Questions about delinquency. *Quality & Quantity* 40: 483-498.
- Becker, R., R. Imhof und M. Raimann, 2007: Kriminalität als rationale Wahlhandlung in der Schweiz: ein Sonderfall? *Schweizerische Zeitschrift für Soziologie* 33: 237-260.
- Becker, R. und R. Günther, 2004: Selektives Antwortverhalten bei Fragen zum delinquenten Handeln. Eine empirische Studie über die Wirksamkeit der „sealed envelope technique“ bei selbst berichteter Delinquenz mit Daten des ALLBUS 2000. *ZUMA-Nachrichten* 54: 39-59.
- Berger, F., 2006: Zur Wirkung unterschiedlicher materieller Incentives in postalischen Befragungen. Ein Literaturbericht. *ZUMA-Nachrichten* 58: 81-100.
- Blau, P.M., 1964: *Exchange and Power in Social Life*. New York: Wiley.
- Blossfeld, H.P. und G. Rohwer, 1995: *Techniques of Event History Modeling – New Approaches to Causal Analysis*. Mahwah, N.J.: Erlbaum.
- Bourdieu, P., 2005: Die Ökonomie der symbolischen Güter (vom Französischen übersetzt). S. 139-155 in: F. Adloff und S. Mau (Hg.): *Vom Geben und Nehmen. Zur Soziologie der Reziprozität*. Frankfurt/New York: Campus.
- Church, A.H., 1993: Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly* 57: 62-79.
- Darven, M., T.H. Rockwood, R. Sherrord und S. Campbell, 2003: Prepaid Monetary Incentives and Data Quality in Face-to-face Interviews. Data from the 1996 Survey of Income and Programm Participation Incentive Experiment. *Public Opinion Quarterly* 67: 139-147.
- De Leeuw, E.D., 1992: Data quality in mail, telephone, and face to face surveys. Amsterdam: TT-Publikaties.
- De Leeuw, E.D., 2001: Reducing Missing Data in Surveys: An Overview of Methods. *Quality & Quantity* 35: 147-160.

- Diekmann, A. und B. Jann, 2001: Anreizformen und Ausschöpfungsquoten bei postalischen Befragungen. Eine Prüfung der Reziprozitätshypothese. ZUMA-Nachrichten 48: 19-27.
- Dillman, D.A., 2000: *Mail and Internet Surveys: The Tailored Design Method*. Second Edition. New York: Wiley.
- Esser, H., 1974: Der Befragte. S. 107-145 in: J. van Koolwijk und M. Wieken-Mayser (Hg.): *Erhebungsmethoden: Die Befragung. Techniken der empirischen Sozialforschung, Band 4*. München: Oldenbourg.
- Esser, H., 1986: Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 314-336.
- Esser, H., 1991: Alltagshandeln und Verstehen. Zum Verhältnis von erklärender und verstehender Soziologie am Beispiel von Alfred Schütz und „Rational Choice“. Tübingen: Mohr.
- Esser, H., 1996: Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48: 1-35.
- Esser, H., 2003: Die Rationalität der Werte. Die Typen des Handelns und das Modell der soziologischen Erklärung. S. 154-188 in: A. Bienfait, G. Albert, C. Wendt und S. Sigmund (Hg.): *Das Weber-Paradigma*. Tübingen: Mohr Siebeck.
- Fox, R.J., M.R. Crask und J. Kim, 1988: Mail Survey Response Rate. A Meta-Analysis of Selected Techniques for Inducing Response. *Public Opinion Quarterly* 52: 467-491.
- Frey, B.S., 1997: *Markt und Motivation. Wie ökonomische Anreize die Arbeits-Moral verdrängen*. München: Franz Vahlen.
- Furse, D.H. und D.W. Stewart, 1982: Monetary Incentives Versus Promised Contribution to Charity: New Evidence on Mail Survey Response. *Journal of Market Research* 19: 375-380.
- Gelb, B.D., 1975: Incentives to Increase Survey Returns: Social Class Considerations. *Journal of Market Research* 12: 107-109.
- Gouldner, A., 1960: The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review* 25: 161-178.
- Goyder, J.C., 1982: Further evidence on factors affecting response rates to mailed questionnaires. *American Sociological Review* 47: 550-553.
- Groves, R.M., E. Singer und A. Corning, 2000: Leverage-Saliency Theory of Survey Participation. Description and an Illustration. *Public Opinion Quarterly* 64: 299-308.
- Harkness, J., P. Mohler, M. Schneid und B. Christoph, 1998: Incentives in Two German Mail Surveys 1996/97 & 1997. S. 201-218 in: A. Koch und R. Porst (Hg.): *Nonresponse in Survey Research. ZUMA-Nachrichten Spezial 4*. Mannheim: ZUMA.
- Hartmann, P.H. und B. Schimpl-Neimanns, 1992: Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 44: 315-340.
- Hartmeier, S. und M.-C. Fontana, 2006: Overreporting: Ein vernachlässigtes Problem und die Schwierigkeit von Gegenmaßnahmen. ZUMA-Nachrichten 58: 50-80.
- Heberlein, T.A. und R.M. Baumgartner, 1978: Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *American Sociological Review* 43: 447-462.
- Hedström, P. und R. Swedberg, 1998: Social mechanisms: An introductory essay. S. 1-31 in: P. Hedström und R. Swedberg (Hg.): *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge: University Press.
- Hill, D.H. und R.J. Willis, 2001: Reducing Panel Attrition. A Search for Effective Policy Instruments. *The Journal of Human Resources* 3: 416-438.
- Hindelang, M.J., T. Hirschi und J.G. Weis, 1979: Correlates of Delinquency: The Illusion of Discrepancy between Self-Report and Official Measures. *American Sociological Review* 44: 995-1014.
- Homans, G.C., 1961: *Social behavior. Its elementary forms*. New York, N.Y.: Harcourt, Brace & World.

- Hubbard, R. und E.L. Little, 1988: Promised Contributions to Charity and Mail Survey Responses. Replication with Extension. *Public Opinion Quarterly* 52: 223-230.
- James, J.M. und R. Bolstein, 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. und R. Bolstein, 1992: Large Monetary Incentives and Their Effects on Mail Survey Response Rates. *Public Opinion Quarterly* 56: 442-453.
- Jobber, D., J. Saunders und V.W. Mitchell, 2004: Prepaid monetary incentive effects on mail survey response. *Journal of Business Research* 57: 21-25.
- Kanuk, L. und C. Berenson, 1975: Mail Surveys and Response Rates: A Literature Review. *Journal of Market Research* 12: 440-453.
- Klein, S. und R. Porst, 2000: Mail Surveys. Ein Literaturbericht. ZUMA-Technischer Bericht 10/2000. Mannheim: ZUMA.
- Koch, A., 1998: Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. *ZUMA-Nachrichten* 42: 66-93.
- Kroneberg, C., 2007: Wertrationalität und das Modell der Frame-Selektion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59: 215-239.
- Krysan, M., H. Schumann, L.J. Scott und P. Beatty, 1994: Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly* 58: 381-399.
- Linsky, A.S., 1975: Stimulating Responses to Mailed Questionnaires: A Review. *Public Opinion Quarterly* 38: 82-101.
- Mauss, M., 1984: Die Gabe. Form und Funktion des Austauschs in archaischen Gesellschaften. Frankfurt am Main: Suhrkamp
- McDaniel, S. und C.P. Rao, 1980: The Effect of Monetary Inducement on Mailed Questionnaire Response Quality. *Journal of Market Research* 17: 265-268.
- Mehlkop, G. und R. Becker, 2007: Zur Wirkung monetärer Anreize auf die Rücklaufquote in postalischen Befragungen zu kriminellen Handlungen. Theoretische Überlegungen und empirische Befunde eines Methodenexperiments. *Methoden – Daten – Analysen* 1: 5-24.
- Nederhof, A.J., 1983: The Effects of Material Incentives in Mail Surveys: Two Studies. *Public Opinion Quarterly* 47: 103-111.
- Neller, K., 2005: Kooperation und Verweigerung: Eine Non-Response-Studie. *ZUMA-Nachrichten* 57: 9-36.
- Porst, R., 1999: Thematik oder Incentives? Zur Erhöhung der Rücklaufquoten bei postalischen Befragungen. *ZUMA-Nachrichten* 23: 72-87.
- Porst, R., 2000: Praxis der Umfrageforschung, 2. Aufl. Stuttgart, Leipzig: Teubner.
- Porst, R., 2001: Wie man die Rücklaufquote bei postalischen Befragungen erhöht. *ZUMA How-to-Reihe* 9. Mannheim: ZUMA.
- Pötter, U. und U. Rendtel, 1993: Über Sinn und Unsinn von Repräsentativitätsstudien. *Allgemeines Statistisches Archiv* 77: 269-280.
- Reuband, K.-H. und J. Blasius, 1996: Face-to-face, telefonische und postalische Befragungen. Ausschöpfungsquoten und Antwortmuster in einer Großstadt-Studie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48: 296-318.
- Riphahn, R.T. und O. Serfling, 2002: Item Non-Response on Income and Wealth Questions. IZA-Working Paper. Bonn: IZA.
- Schewe, C.D. und N.G. Cournoyer, 1976: Prepaid vs. Promised Monetary Incentives to Questionnaire Response: Further Evidence. *Public Opinion Quarterly* 40: 105-107.
- Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske+Budrich.
- Schräpler, J.-P., 2001: Respondent Behaviour in Panel Studies. A Case Study of the German Socio-Economic Panel (GSOEP). DIW-Discussion Paper 244. Berlin: DIW.

- Simon, H.A., 1993: Homo rationalis. Die Vernunft im menschlichen Leben. Frankfurt am Main: Campus.
- Singer, E., 1998: Incentives for Survey Participation: Research on Intended and Unintended Consequences. ZUMA-Nachrichten 22: 7-29.
- Singer, E., J. van Hoeyk und M.P. Maher, 1998: Does the Payment of Incentives Create Expectation Effects? Public Opinion Quarterly 62: 152-164.
- Stadtmüller, S. und R. Porst, 2005: Zum Einsatz von Incentives bei postalischen Befragungen. ZUMA How-to-Reihe 14. Mannheim: ZUMA.
- Sudman, S. und N.M. Bradburn, 1982: Asking Questions: A practical guide to questionnaire design. San Francisco: Jossey-Bass.
- Sudman, S. und N.M. Bradburn, 1974: Response Effects in Surveys: A Review and Synthesis. Chicago: Aldine.
- Warriner, K., J. Goyder, H. Gjersten, P. Hohner, und K. McSpurren, 1996: Charities, No; Lotteries, No; Cash, Yes. Main Effects and Interactions In A Canadian Incentives Experiment. Public Opinion Quarterly 60: 542-562.
- Yammarino, F.J., S.J. Skinner und T.L. Childers, 1991: Understanding Mail Survey Response Behavior. Public Opinion Quarterly 55: 613-639.
- Yu, J. und H. Cooper, 1983: A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. Journal of Market Research 20: 36-44.

Korrespondenzadressen:

Rolf Becker
Universität Bern
Institut für Erziehungswissenschaften
Abteilung Bildungssoziologie
Muesmattstraße 27
CH-3012 Bern
rolf.becker@edu.unibe.ch

Regula Imhof
Universität Bern
Institut für Erziehungswissenschaften
Abteilung Bildungssoziologie
Muesmattstraße 27
CH-3012 Bern
regula.imhof@edu.unibe.ch

Guido Mehlikop
Technische Universität Dresden
Institut für Soziologie
Mommsenstraße 13
01069 Dresden
guido.mehlikop@mailbox.tu-dresden.de

Die Kombination von Mobilfunk- und Festnetzstichproben

Combining mobile phone and landline phone samples

Eine praktische Anwendung des Dual-Frame-Ansatzes

A practical application of the dual frame approach

Stefan Hunsicker und Yvonne Schroth

Zusammenfassung

Schwierige Erreichbarkeit sehr mobiler Personen, steigende Verweigerungsraten durch die Zunahme von Marketing- und Werbeanrufen sowie der gestiegene Anteil an exklusiven Mobilfunknutzern sind einige der Gründe, die den Schluss nahelegen, dass repräsentative Bevölkerungsumfragen zukünftig nur noch mit einem kombinierten Stichprobendesign gewährleistet werden könnten. Gegenwärtig wird der Dual-Frame-Ansatz von Festnetz- und Mobilfunkstichproben im Hinblick auf seine Praxistauglichkeit getestet. In diesem Kontext untersucht der vorliegende Beitrag anhand aktueller Umfragen der Forschungsgruppe Wahlen, welche Auswirkungen sich auf die Zusammensetzung der Stichprobe ergeben und wie sich die Verteilungen politisch relevanter Variablen verändern, wenn Festnetzstichproben in Kombination mit Mobilfunkstichproben verwendet werden.

Abstract

The difficulty in contacting very mobile people, increasing non-response as a result of the flood of marketing and cold calls, and the growing number of people who use mobile phones exclusively are all reasons to argue in favor of using a sampling structure that combines mobile phone and landline phone sample designs to ensure the quality of telephone surveys. Currently, a dual-frame concept of landline and mobile phone samples is being put to the test. In this context, this article examines the effects on the distribution of politically relevant variables in current surveys of the Forschungsgruppe Wahlen, where mobile phone numbers are included in random telephone samples.

1 Einleitung und Fragestellung¹

Es geht wohl kein Weg daran vorbei: man muss sich einer grundsätzlichen Auseinandersetzung über die Integration von Mobilfunknummern in telefonischen Umfragen der Meinungs- und Marktforschung stellen. Die Gründe dafür sind keinesfalls trivial und den Instituten, die über Festnetzstichproben CATI-Befragungen durchführen, wohl bekannt. Probleme der CATI-Befragung liegen einerseits in der schwierigen Erreichbarkeit von sehr mobilen Personen über ihren Festnetzanschluss und andererseits im Anteil derjenigen, die man über Festnetzstichproben gar nicht mehr erreicht. Dieser Anteil an Personen, die durch ihre exklusive Mobilfunknutzung systematisch aus repräsentativen Telefonstichproben ausgeschlossen sind, macht ungefähr 5–7% der Bevölkerung aus, was noch als relativ unproblematisch zu werten ist.² Ein stetig steigender Anteil an exklusiven Mobilfunknutzern würde die Repräsentativität von Bevölkerungsumfragen, deren Ergebnisse ausschließlich durch das Festnetz gewonnen werden, immer mehr infrage stellen, und die Branche stünde vor einem Problem.

Eine zentrale Problematik der Integration von Mobilfunknummern in Festnetzstichproben liegt darin, dass Handynummern sich regional nicht verorten lassen, also eine entsprechende Schichtung nicht möglich ist. Zur Verbesserung der Erreichbarkeit bzw. zur Erreichbarkeit der exklusiven Handynutzer über ein Screeningverfahren könnte man sich damit behelfen, die ungefähr zwei Millionen im Telefonbuch eingetragenen und nicht als gewerblich gekennzeichneten Handynummern zu nutzen. Diese könnten durch die Gemeinde der Eintragung in die regionale Schichtung mit einbezogen werden. Dieses Vorgehen scheint aber nur dann sinnvoll, wenn sich die Gruppe der Handynutzer, welche ihre Mobilfunknummer im Telefonbuch eingetragen haben, auch nicht von den restlichen Mobilfunknutzern unterscheidet. Vor allem dürften sich die beiden Gruppen nicht in Bezug auf sozialstrukturelle Merkmale, wie z. B. Bildungsstatus, unterscheiden, um einen Befragungsbias mit damit korrelierenden politischen und sozialen Einstellungen zu vermeiden. Offen ist auch die Frage, ob die Integration von Mobilfunknummern in Festnetzstichproben so zu bewerkstelligen ist, dass wir ohne methodische Qualitätsverluste zu einer Verbesserung der Erreichbarkeit bestimmter, über das Festnetz

1 Wir danken den Herausgebern und den anonymen Gutachtern für wertvolle Hinweise.

2 In einer aktuellen face to face (f2f)-Befragung im Auftrag der Forschungsgruppe Wahlen Ende 2006 liegt der Anteil der exklusiven Mobilfunknutzer bei 5,7% der befragten Personen. Ähnliche Werte finden sich bei Glemser (2007: 11).

schwierig zu erreichender Personengruppen gelangen und auch exklusive Handynutzer prinzipiell erreichen und befragen können.³

Zusammenfassend stehen in unserem Beitrag drei Fragen im Mittelpunkt:

1. Unterscheiden sich die Subpopulationen von eingetragenen und nicht eingetragenen Mobilfunknutzern nach regionalen, sozialstrukturellen und politisch relevanten Merkmalen?
2. Welche Bevölkerungsgruppen verbergen sich hinter den exklusiven Mobilfunknutzern? Wer wird in Festnetzumfragen also unterrepräsentiert und welche systematischen Verzerrungen werden in gegenwärtigen Stichproben in Kauf genommen, die exklusive Mobilfunknutzer nicht integrieren? In welchen Merkmalen unterscheiden sich die beiden Subpopulationen der Mobil- und Festnetznutzer gegenüber reinen Mobilfunknutzern?
3. Als vermeintlicher Königsweg gilt in der aktuellen Diskussion ein kombiniertes Stichprobendesign, bestehend aus einer Festnetzstichprobe (inkl. Nur-Festnetznutzer) und einer Mobilfunkstichprobe (inkl. Nur-Mobilnutzer). Hier gehen wir der Frage nach, wie sich die Umfragewerte politisch-inhaltlicher Variablen (Sonntagsfrage, politisches Interesse, Wahlbeteiligung) innerhalb eines solchen Designs gegenüber einer reinen Festnetzstichprobe verändern.

Die Relevanz dieser und anderer Fragen lässt sich nicht zuletzt in einer vom ZUMA ins Leben gerufenen Arbeitsgruppe „Mobilsample“⁴ sowie einer im November 2006 stattgefundenen ZUMA-Tagung und der daraus entstandenen Publikation erahnen (vgl. Gabler/Häder 2007).

2 Datenbasis

Die Analyse basiert auf vier unterschiedlichen Stichproben: eine Einschaltung in eine face to face (f2f)-Stichprobe nach dem ADM-Design, zwei reine Mobilfunkstichproben (mobein, mobgen) und eine reine Festnetzstichprobe (PB).

Bei der f2f-Studie handelt es sich um eine Einschaltung in den USUMA face-to-face-Bus nach ADM-Design. Als Festnetzstichprobe wurde das Politbarometer vom Dezember 2006 verwendet (vgl. zum Design Schroth 2003; Schulte 1997).

- 3 Die Problematisierung weiterer Aspekte eines Auswahlrahmens für Mobilfunknummern, wie Gewichtung, Kontaktsituation und Roaming sowie Mode effects, können wir in diesem Rahmen nicht leisten. Diese Aspekte werden ausführlich in dem DFG-Projekt des ZUMA „Telefonbefragung in der Allgemeinbevölkerung über das Mobilfunknetz“ vertieft (vgl. Schneiderat/Häder 2007).
- 4 Vgl. Mitteilung über die Gründung der Arbeitsgruppe MOBILSAMPLE. ZUMA-Nachrichten 56, 2005: 111ff.

Tabelle 1 Datenbasis⁵

Studie	Erhebungszeitraum	Grundgesamtheit	Stichprobe	Wahlberechtigte
mobein	05.12 - 07.12.2006	Wohnbevölkerung ab 14 Jahren	n = 1572	n = 1477
mobgen	05.12 - 07.12.2006	Wohnbevölkerung ab 14 Jahren	n = 1538	n = 1268
f2f	19.11 - 18.12.2006	Wohnbevölkerung ab 14 Jahren	n = 4995	n = 4695
PB	12.12 - 14.12.2006	Wahlberechtigte	n = 1633	n = 1633

Die beiden Mobilfunkstichproben unterscheiden sich dadurch, dass es sich bei ersterer um eine Auswahl von im Telefonbuch eingetragenen Mobilfunknummern (mobein) handelt, während die zweite aus einem Auswahlrahmen von frei generierten Handynummern gezogen wurde (mobgen). Der Auswahlrahmen eingetragener Mobilfunkanschlüsse, der eine regionale Verortung entsprechend der Festnetzstichproben ermöglicht, umfasst etwa zwei Mio. (als privat kategorisierte) Einträge. Die Konstruktion des zweiten Auswahlrahmens erfolgte über die Generierung von Ziffernfolgen innerhalb der bestehenden Netz-Vorwahlen und umfasst ca. 175 Mio. Mobilfunknummern. Generiert wurden pro Netz alle Nummern in 100.000er Blöcken, die mindestens einen als privat gekennzeichneten Eintrag aufweisen. Zudem erfolgt ein Abgleich mit der sog. Gajek-Liste⁶. Die Nutzung dieser Informationsquelle führt einerseits zum Ausschluss von Blöcken mit Sondereigenschaften (Mobilboxabfragen, Sonderzwecke etc.), andererseits zur Ergänzung der Auswahlgrundlage mit neu eröffneten 100.000er Blöcken. So stammen 34,7 Mio. Nummern aus Blöcken ohne jeglichen Eintrag, davon 20 Mio. in den zwei neuen 8-stelligen Blöcken 0151/5... und 0157/7.

Bei den Mobilfunkstudien wurden an 3 Feldtagen 18.845 Nummern verwendet, 5.002 eingetragene und 13.843 generierte Nummern aus dem FGW-Mobilfunkpool. Die Studien wurde an allen Tagen im ‚predictive mode‘ durchgeführt. Realisiert wurden insgesamt 3.110 Interviews, davon 1.572 aus eingetragenen Nummern und 1.538 aus dem FGW-Mobilfunkpool. Gefragt wurde nach der Handyausstattung der befragten Person und dem dazugehörenden Vertrag, ebenso nach der Telefonausstattung des Haushalts. Befragten über 18 Jahren wurden noch einige Fragen zu politischen Einstellungen sowie die Wahlabsichtsfrage gestellt und Merkmale der Standarddemographie erhoben.

5 In der realisierten Stichprobe aus dem FGW-Mobilfunkpool (mobgen) finden sich 62 Nummern (Wahlberechtigte: 58), die im Telefonbuch eingetragen waren.

6 <http://www.gajek.de> sowie Heckel (2007). Dank an C. Heckel (BIK Aschpurwis+Behrens) für den Hinweis zur Gajek-Liste, die von G. Schneiderat im Internet entdeckt wurde.

Im Folgenden werden einige Informationen zur Regionalverteilung und zur Ausschöpfung der Stichprobe aus dem FGW-Mobilfunkpool (mobgen) berichtet, die in Abschnitt 4 mit dem Politbarometer nach dem Dual-Frame-Ansatz zusammengewichtet wird. Die Verteilung der Interviews ist, obwohl ohne jegliche Schichtung durchgeführt, durchaus zufriedenstellend, wie Tabelle 2 zeigt.

Tabelle 2 Regionale Verteilung der Interviews aus der Mobilfunkstudie (mobgen)

Regionale Verteilung der Interviews aus der Mobilfunkstudie der generierten Nummern (Wohnbevölkerung ab 14 Jahren) in %	Bevölkerungsfortschreibung 31.12.2004	
1 Schleswig-Holstein	3,0	3,4
2 Hamburg	2,5	2,1
3 Niedersachsen	7,9	9,6
4 Bremen	0,8	0,8
5 Nordrhein-Westfalen	21,9	21,7
6 Hessen	7,1	7,4
7 Rheinland-Pfalz	4,4	4,9
8 Baden-Württemberg	11,2	12,8
9 Bayern	17,2	14,9
10 Saarland	1,2	1,3
11 Berlin (ehemaliger Westteil)	3,1	2,5
12 Berlin (ehemaliger Ostteil)	2,2	1,7
13 Brandenburg	3,1	3,2
14 Mecklenburg-Vorpommern	2,3	2,2
15 Sachsen	6,5	5,4
16 Sachsen-Anhalt	3,2	3,2
17 Thüringen	2,4	3,0
Gesamt	n = 1538	71,49 Mio.

Datenbasis: designgewichtete Daten

Die Ausschöpfung der Stichprobe aus dem FGW-Mobilfunkpool (mobgen) liegt bei 55% (vgl. Tabelle 3). Die Berechnung der Ausschöpfung ist mit einigen Schwierigkeiten behaftet. Bei Nichtkontakt des Teilnehmers ist nicht immer eindeutig entscheidbar, ob es sich um einen nicht vergebenen Anschluss handelt (nicht zur Stichprobe gehörend, d. h. neutraler Ausfall) oder der Teilnehmer vorübergehend nicht erreichbar ist (kein Kontakt) (vgl. Häder/Glemser 2006: 169).

Tabelle 3 Ausschöpfung der Stichprobe aus dem FGW-Mobilfunkpool (mobgen)

1	Brutto-Ausgangsstichprobe	13843	
2	davon nicht zur Stichprobe	8305	60,0% (von 1)
3	Bereinigte Bruttostichprobe	5538	
4	davon kein Kontakt	2748	49,6% (von 3)
5	Brutto mit Kontakt	2790	100%
	Ausfälle:		
6	Zielperson verweigert	920	33,0%
7	Abbruch	9	1,0%
8	Sonstiger Ausfall	303	10,9%
9	Realisierte Interviews	1538	55,1% (von 5)

In einem Pretest (4.869 Nummern, 650 Interviews) wurden alle vom Dialer im ‚predictive mode‘ nicht vorgelegten Nummern von den Interviewern angehört (‚preview mode‘) und vercodet. Dabei erwiesen sich 1.949 von 2.189 Nummern, das heißt 89%, als ‚Anfragen der Provider/Netzanbieter: Nummer nicht vergeben‘. Für Befragungen, die sich auf Grundgesamtheiten beschränken, welche nicht der gesamten Wohnbevölkerung entsprechen, erweist sich zudem die Einschätzung der Kategorie ‚Zielperson verweigert‘ (sofortiges Auflegen, keine Zeit, kein Interesse, keine Auskunft) als problematisch. Es ist zum Beispiel nicht zu entscheiden, ob ein Teilnehmer, der sofort auflegt, zur Grundgesamtheit der Wahlberechtigten gehört. Die Ausschöpfungsquoten werden in diesem Fall unterschätzt.

3 Empirische Ergebnisse

3.1 Haushalte und Personen nach Telefonanschluss

Um über Festnetzstichproben repräsentative Telefonumfragen zu erheben, muss stets auf Veränderungen in der Telefonausstattung von Haushalten reagiert werden. Informationen, wie die Nutzung der vielfältigen Möglichkeiten der Telekommunikation verteilt ist, können aktuell aus der EVS von 2003 (vgl. Statistisches Bundesamt 2003; Heckel 2007: 33ff.) und aus f2f-Befragungen gewonnen werden. Bei f2f-Befragungen ist zu vermuten, dass der Anteil der Mobilfunknutzer tendenziell etwas

unterschätzt wird (vgl. Häder/Glemser 2006: 164). Die nachfolgenden Tabellen zeigen die Ausstattung der Haushalte bzw. der Personen sowie der Wahlberechtigten mit Telefonanschlüssen auf der Basis der Daten der f2f-Stichprobe.

Tabelle 4 Telefonausstattung

	Dezember 2006		
	West	Ost	Gesamt
Haushalte in %			
Festnetzanschluss	92,9	86,0	91,5
nur Mobilfunk	5,9	11,2	7,0
Telefonhaushalte insgesamt	98,8	97,2	98,5
kein Telefonbesitz	1,2	2,8	1,5
Personen in %^a			
Festnetzanschluss	94,6	87,6	93,3
nur Mobilfunk	4,6	10,4	5,7
Telefonhaushalte insgesamt	99,2	98,0	99,0
kein Telefonbesitz	0,8	2,0	1,0
Wahlberechtigte in %^b			
Festnetzanschluss	95,0	87,7	93,6
nur Mobilfunk	4,1	10,3	5,4
Telefonhaushalte insgesamt	99,2	98,0	99,0
kein Telefonbesitz	0,8	2,0	1,0

Datenbasis: USUMA-f2f-Bus (West inkl. Westberlin, Ost inkl. Ostberlin).

^a *n = 4995, korrekturegewichtet nach Alter und Geschlecht.*

^b *n = 4645, korrekturegewichtet nach Alter und Geschlecht.*

Problematisch im Hinblick auf reine Festnetzstichproben sind die 7% der Haushalte bzw. rund 6% der Personen, die nur noch über Mobilfunk zu erreichen sind. 91% der Haushalte sind aber immer noch über das Festnetz erreichbar (71% über einen analogen Anschluss, 19% über ISDN, 5% über VoIP, 3% durch Homezone und 1% über Kabel).⁷ Im Osten liegt sowohl der Anteil der Haushalte ohne Telefonbesitz (3%) als auch der Anteil der reinen Mobilfunkhaushalte (11%) deutlich höher als im Westen. Der Anteil von 6% Mobilonlys hat sich von der Größenordnung her in den letzten Jahren kaum verändert (vgl. Glemser 2007: 11), d. h. die enormen Wachstumsraten auf dem Mobilfunksektor haben keineswegs zu einem massenhaften Verschwinden

7 Bei der Erreichbarkeit über das ‚Festnetz‘ sind Mehrfachausstattungen möglich.

der potentiellen telefonischen Erreichbarkeit über Telefonanschlüsse mit Ortsnetz-kennziffern geführt.⁸ Der Anteil der Mobilonlys an den Wahlberechtigten ist geringfügig niedriger als deren Anteil an der Wohnbevölkerung ab 14 Jahren. Die nachfolgenden Analysen beschränken sich auf die wahlberechtigte Bevölkerung, da unser hauptsächliches Erkenntnisinteresse darin besteht, die Veränderungen politisch relevanter Variablen durch die Berücksichtigung von Mobilfunkteilnehmern zu untersuchen.

3.2 Eingetragene und nicht eingetragene Mobilfunkteilnehmer

Zwischen den befragten Wahlberechtigten mit einem Eintrag ihrer Mobilfunknummer ins Telefonbuch und Befragten ohne Eintrag gibt es zunächst deutliche Unterschiede hinsichtlich der Verträge. Lediglich 5% der eingetragenen Teilnehmer besitzen ein Prepaid-Handy, 12% einen Homezone-Vertrag und 83% einen festen Vertrag. Bei den Befragten mit generierter Nummer beträgt der Anteil der Prepaid-Kunden dagegen 34%, 7% nutzen Homezone und 59% haben einen festen Vertrag.

Die Ergebnisse einer binären Logit-Analyse zeigen, dass die wichtigsten Prädiktoren eines Mobilfunkeintrags die Bildung, das Geschlecht und der Status der Erwerbstätigkeit sind (Modell 1 in Tabelle 5). Im Gegensatz zu den Befragten ohne Eintrag handelt es sich bei denjenigen mit Eintrag eher um Männer, Befragte mit höherem Bildungsniveau sowie Erwerbstätige im Vergleich zu Befragten, die studieren, sich in Ausbildung befinden oder arbeitslos sind. Personen mit Eintrag weisen ein signifikant höheres politisches Interesse auf. West-Ost-Unterschiede sind keine zu beobachten. Auch bei den wahlberechtigten Mobilonlys (Modell 2 in Tabelle 5) gibt es zwischen eingetragenen und generierten Nummern prinzipiell ähnliche Unterschiede. Der geschlechtsspezifische Unterschied ist allerdings nicht mehr zu beobachten. Die Differenzen im Hinblick auf das politische Interesse existieren nach wie vor.

Insgesamt unterstreichen die systematischen Unterschiede, sowohl bei den sozialstrukturellen als auch den inhaltlichen Variablen, dass ein Mobilfunk-Auswahlrahmen sich nicht auf den eingetragenen Rufnummernraum beschränken kann, auch nicht zu dem Zweck, nur Mobilonlys als Ergänzung für Festnetzstichproben daraus zu rekrutieren.

8 Im TNS Infratest f2f-Bus werden für 2006 7,6% Mobilfunkhaushalte, 90,2% ‚Festnetz‘-Haushalte und 2,2% Haushalte ohne Telefonbesitz ermittelt.

Tabelle 5 Binäre Logit-Analyse für den Eintrag der Mobilfunknummer ins Telefonbuch^a

	Modell 1 Wahlberechtigte			Modell 2 Wahlberechtigte Mobilonlys		
	B	Wald	Sig.	B	Wald	Sig.
Konstanter Term	-0,310	1,261	0,262	-1,277	2,374	0,123
Gebiet (<i>Ref. Ost</i>)	0,067	0,374	0,541	-0,240	0,633	0,426
Ortsgröße						
bis 20.000 Einw.	0,321	9,100	0,003	1,526	17,909	0,000
bis 100.000 Einw.	0,128	1,049	0,306	1,203	9,668	0,002
<i>Ref. mehr als 100.000 Einw.</i>	0	.	.	0	.	.
Geschlecht (<i>Ref. Frau</i>)	0,543	32,243	0,000	-0,119	0,147	0,702
Alter						
18 bis 29 Jahre	-0,283	3,453	0,063	-0,632	1,728	0,189
30 bis 39 Jahre	-0,203	2,021	0,155	-0,871	3,091	0,079
40 bis 49 Jahre	-0,141	0,954	0,329	-0,877	2,717	0,099
<i>Ref. 50 Jahre und älter</i>	0	.	.	0	.	.
Bildung						
ohne/Volks-/Hauptschulab.	-0,611	22,036	0,000	-0,957	5,420	0,020
mittlere Reife/POS	-0,468	19,208	0,000	-0,851	5,584	0,018
<i>Ref. Abitur</i>	0	.	.	0	.	.
Erwerbstatus						
arbeitslos	-0,700	9,357	0,002	-0,970	3,264	0,071
Rentner	0,250	1,544	0,214	-0,359	,301	0,584
in Ausbildung/Hochschule	-0,486	7,715	0,005	-1,290	6,298	0,012
nicht erwerbstätig/Haushalt	-0,282	2,058	0,151	-0,425	0,318	0,573
<i>Ref. erwerbstätig</i>	0	.	.	0	.	.
Eigene wirtschaftl. Lage ^b	0,066	1,016	0,313	0,253	1,608	0,205
Politisches Interesse ^c	0,128	8,493	0,004	0,338	6,557	0,010
Fallzahl	N = 2687			N = 338		
Pseudo R ²	4,7%			13,8%		
Basis: designgewichtete Daten						

^a Abhängige Variable: 0 Nummer generiert, 1 Nummer eingetragen.^b Eigene wirtschaftliche Lage: 1 schlecht, 2 teils/teils, 3 gut.^c politisches Interesse: 1 gar nicht, 2 kaum, 3 etwas, 4 stark, 5 sehr stark.

3.3 Exklusive Mobilfunknutzer

Unterscheiden sich die exklusiven Mobilfunknutzer hinsichtlich ihrer politischen Einstellungen und Bewertungen von den Wahlberechtigten, die man prinzipiell auf einem Festnetzanschluss erreichen kann? Um Rückschlüsse auf Personenkreise ziehen zu können, deren politische Überzeugung wir in Umfragen mit Festnetzstichproben bisher nicht abbilden können, wird zwischen folgenden Nutzungstypen unterschieden: Mobilonlys (1), Mobilfunknutzer mit analogem Telefonanschluss im Haushalt (2), mit ISDN-Anschluss (3) und sonstigem Anschluss wie Kabel-Telefonie oder VoIP (4) sowie Mobilfunkteilnehmer mit ‚Home-Zone‘-Vertrag (5). Unter den Befragten mit eingetragenen Mobilfunknummern haben wir 149 Personen mit exklusiver Mobilfunknutzung und unter den Befragten aus dem FGW-Mobilfunkpool sind es 262.

Mit Hilfe von Korrespondenzanalysen⁹ suchen wir nach Zusammenhängen zwischen Nutzungstyp, politischen Einstellungen und sozialstrukturellen Hintergrundvariablen. Ziel ist es, Unterschiede der Nutzungstypen bezüglich ihrer politischen Motivationen und ihrer sozialen Herkunft aufzuzeigen.

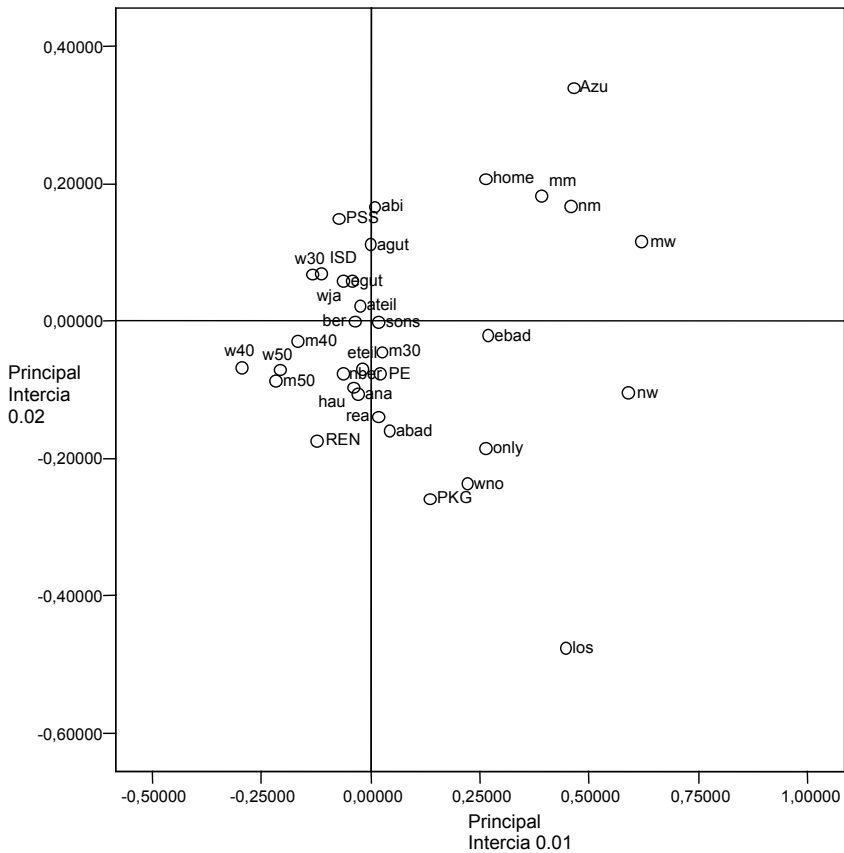
Das erste Modell umfasst die Befragten mit eingetragenen Mobilfunknummern (mobein). Die erste Achse (die Abszisse) erklärt 50% der Varianz des gesamten Modells, die zweite Achse (die Ordinate) erklärt weitere 35%. Auf der ersten Achse lassen sich auf der rechten Seite die Alterskategorien bis 29 Jahre und auf der linken Seite diejenigen von 30 Jahre und älter projizieren, so dass diese Achse als Altersdimension bezeichnet werden kann. Mit dieser Altersdimension korrelieren die Kategorien der exklusiven Handynutzer und die Gruppe der Home-Zone-Nutzer, die beide auf der rechten Seite projiziert wurden, sowie die Gruppe mit ISDN-Anschluss, die sich auf der linken Seite befindet.

Mit Hilfe des numerischen Ergebnisses kann die Korrelation der Merkmalsausprägungen mit der ersten Achse näher betrachtet werden.¹⁰

9 Das Verfahren der Korrespondenzanalyse ist ein Skalierungsverfahren, das kategoriale und höher skalierte Merkmale auf latente Dimensionen reduziert, welche wie Faktoren einer Hauptkomponentenanalyse zu interpretieren und zu benennen sind. Die Korrespondenzanalyse stellt, ähnlich der Clusteranalyse, ein Klassifikationsinstrument dar (vgl. Blasius 2001). Die (einfachen und symmetrischen) Korrespondenzanalysen in diesem Beitrag wurden mit dem Programm SIMCA (Version 2.0) von M.J. Greenacre berechnet.

10 Zur Interpretation der numerischen Ergebnisse der Korrespondenzanalyse greifen wir zurück auf die Hinweise von Blasius (1988, 2001).

Abbildung 1 Modell I: Typen der Telefonausstattung
(eingetragene Mobilfunknummern)



Variablenlegende:

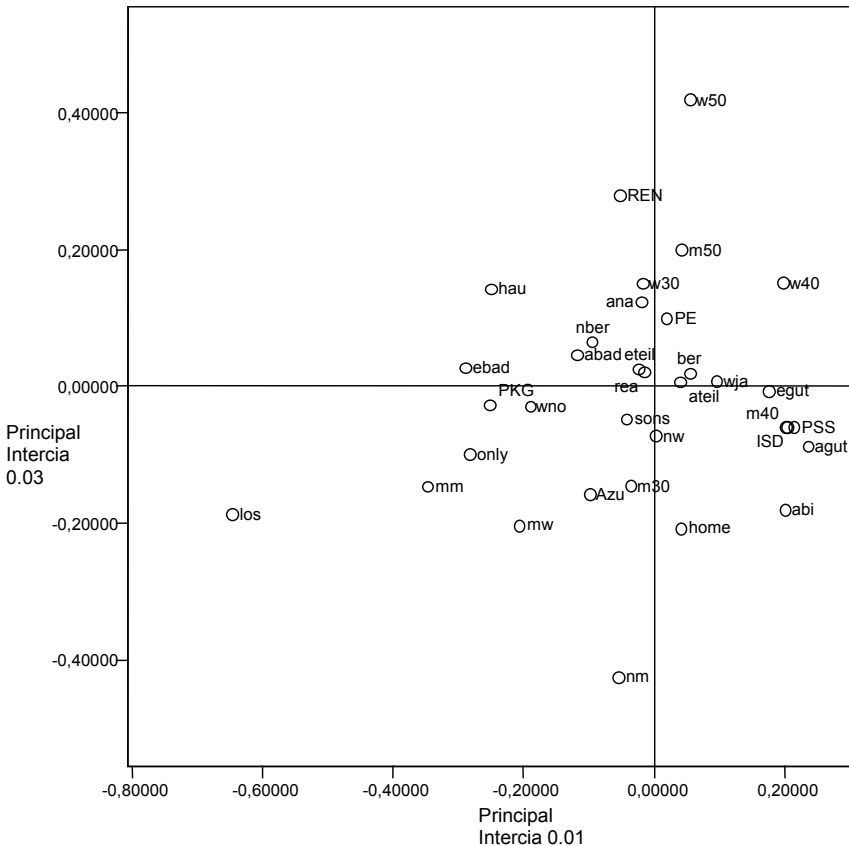
only: exklusive Mobilfunknutzer; **home:** Home-Zone Nutzer; **ana:** Befragte mit analogem Telefonanschluss im Haushalt; **ISD:** Befragte mit ISDN-Telefonanschluss im Haushalt; **sons:** Befragte, die über Kabelanschluss oder Internet telefonieren.

PSS: sehr starkes oder starkes Politikinteresse; **PE:** etwas politisches Interesse; **PKG:** kaum oder gar kein Politikinteresse; **egut:** eigene wirtschaftliche Lage als gut eingeschätzt; **etei:** eigene wirtschaftliche Lage teils gut/teils schlecht; **ebad:** eigene wirtschaftliche Lage als schlecht eingeschätzt; **agut:** allgemeine wirtschaftliche Lage als gut eingeschätzt; **atei:** allgemein wirtschaftliche Lage teils gut/teils schlecht; **abad:** allgemeine wirtschaftliche Lage als schlecht eingeschätzt; **wja:** würde zur Wahl gehen; **wno:** würde nicht zur Wahl gehen; **mm:** 18-24 Jahre männlich; **mw:** 18-24J. weiblich; **nm:** 25-29J. männl.; **nw:** 25-29J. weibl.; **m30:** 30-39J. männl.; **w30:** 30-39J. weibl.; **m40:** 40-49J. männl.; **w40:** 40-49J. weibl.; **m50:** 50J. und älter männl.; **w50:** 50J. und älter weibl.; **hau:** Hauptschulabschluss; **rea:** Realschulabschluss; **abi:** Fachhochschulreife/Abitur; **ber:** Voll-/Teilzeit berufstätig; **los:** arbeitslos; **REN:** Renter/Pension; **Azu:** in Ausbildung/Hochschule; **nber:** nicht berufstätig/Hausfrau/-mann.

Hinter der Gruppe der exklusiven Handynutzer (only) und der Home-Zone-Nutzer mit Telefonbucheintrag (home) verbergen sich demnach überdurchschnittlich häufig Frauen und Männer von 18 bis 29 Jahren, die sich gerade in der Ausbildung oder im Studium befinden und ihre eigene wirtschaftliche Lage als nicht günstig bezeichnen. Die Gruppe der Arbeitslosen korreliert lediglich gering mit (0,37) mit der ersten Achse. Die Befragten mit ISDN-Anschluss (ISD) sind im Gegensatz dazu überdurchschnittlich häufig berufstätig und bei den Frauen älter als 29 Jahre und bei den Männern älter als 40 Jahre. Die Kategorie der Männer von 30-39 Jahre korreliert mit keiner der Achsen. Die zweite Achse lässt sich als Bildungsachse beschreiben. Auf ihr lassen sich im oberen Teil die Personen mit Abitur projizieren, ebenso die Kategorien derer, die sich stark für Politik interessieren, die allgemeine wirtschaftliche Lage als gut bezeichnen sowie diejenigen, die eine positive Wahlteilnahme bescheinigen, wenn denn am nächsten Sonntag Wahl wäre. Negativ zu diesen Kategorien korrelieren die Gruppen mit schlechter Einschätzung der wirtschaftlichen Lage, die Kategorie der Nichtwähler bei einer imaginären Wahl sowie die Gruppe mit mittlerem Bildungsabschluss. Diese Personengruppen haben in Relation zu den anderen Kategorien häufiger einen analogen Telefonanschluss (ana). Auf einer dritten Dimension mit weiteren 11,5% der Varianzaufklärung, die graphisch nicht dargestellt ist, korrelieren lediglich die Ausprägungen ‚Rentner‘ und ‚nicht erwerbstätig‘. Die Kategorie mit sonstiger Telefonausstattung im Haushalt (sons) korreliert mit keiner der ersten drei Dimensionen. Zusammenfassend lässt sich der Raum der Telefonausstattung und Nutzung durch die Dimensionen Alter und Bildung mit Politikinteresse interpretieren. Die erste Achse beschreibt die altersspezifische Ausstattung von ISDN-Anschlüssen im Gegensatz zur Home-Zone-Nutzung und zur exklusiven Mobilfunknutzung. Mit der zweiten Bildungsachse korreliert lediglich die Gruppe mit analogen Telefonanschlüssen im Haushalt.

Während sich die unterschiedlichen Vorkommenshäufigkeiten der sozialstrukturellen und soziodemographischen Variablen Schulbildung, berufliche Tätigkeit sowie Alter und Geschlecht in den fünf Gruppen von eingetragenen Telefonnutzern zu 85% Prozent über die beiden Dimensionen ‚Alter‘ und ‚Bildung‘ interpretieren lassen, betrachten wir nun das Modell der Befragten, deren Mobilfunknummern nicht im Telefonbuch eingetragen, also generiert sind (Abbildung 2). Hier haben sich die Achsen ‚umgedreht‘. Die Abszisse, die mit 53,5% allein über die Hälfte der Varianz erklärt, lässt sich als Bildungsachse interpretieren und die Ordinate als Altersdimension, die weitere 24,8% der Gesamtvarianz des Modells erklärt. Im Gegensatz zum ersten Modell der Befragten mit im Telefonbuch eingetragenen Mobilfunknummern korreliert hier die Gruppe der exklusiven Mobilfunknutzer stark negativ mit der Bildungsdimension.

Abbildung 2 Modell II: Typen der Telefonausstattung
(generierte Mobilfunknummern)



Variablenlegende:

only: exklusive Mobilfunknutzer; **home:** Home-Zone Nutzer; **ana:** Befragte mit analogem Telefonanschluss im Haushalt; **ISD:** Befragte mit ISDN-Telefonanschluss im Haushalt; **sons:** Befragte, die über Kabelanschluss oder Internet telefonieren.

PSS: sehr starkes oder starkes Politikinteresse; **PE:** etwas politisches Interesse; **PKG:** kaum oder gar kein Politikinteresse; **egut:** eigene wirtschaftliche Lage als gut eingeschätzt; **etei:** eigene wirtschaftliche Lage teils gut/teils schlecht; **ebad:** eigene wirtschaftliche Lage als schlecht eingeschätzt; **agut:** allgemeine wirtschaftliche Lage als gut eingeschätzt; **atei:** allgemein wirtschaftliche Lage teils gut/teils schlecht; **abad:** allgemeine wirtschaftliche Lage als schlecht eingeschätzt; **wja:** würde zur Wahl gehen; **wno:** würde nicht zur Wahl gehen; **mm:** 18-24 Jahre männlich; **mw:** 18-24J. weiblich; **nm:** 25-29J. männl.; **nw:** 25-29J. weibl.; **m30:** 30-39J. männl.; **w30:** 30-39J. weibl.; **m40:** 40-49J. männl.; **w40:** 40-49J. weibl.; **m50:** 50J. und älter männl.; **w50:** 50J. und älter weibl.; **hau:** Hauptschulabschluss; **rea:** Realschulabschluss; **abi:** Fachhochschulreife/Abitur; **ber:** Voll-/Teilzeit berufstätig; **los:** arbeitslos; **REN:** Renter/Pensionär; **Azu:** in Ausbildung/Hochschule; **nber:** nicht berufstätig/Hausfrau/-mann.

Im relativen Vergleich zu den anderen Ausprägungen verfügen die exklusiven Handynutzer überdurchschnittlich häufig über einen Hauptschulabschluss und geben überdurchschnittlich oft an, arbeitslos zu sein. Es sind meistens junge Frauen oder Männer zwischen 18 und 24 Jahren, die relativ oft über kaum bzw. gar kein Politikinteresse verfügen und sowohl die eigene als auch die allgemeine wirtschaftliche Lage als schlecht einschätzen. Im Vergleich dazu befanden sich die exklusiven Mobilfunknutzer mit eingetragener Handynummer überdurchschnittlich häufig in der Ausbildung bzw. im Studium und hatten eine hohe Ladung auf der Alters- und nicht auf der Bildungsdimension.

Stark positiv mit der Bildungsachse korreliert die Gruppe der ISDN-Nutzer, die auf der rechten Seite projiziert ist. Diese Kategorie lässt sich dadurch beschreiben, dass sie im Gegensatz zu den exklusiven Mobilfunknutzern häufiger über einen hohen Bildungsabschluss verfügt, sich sehr stark für Politik interessiert und auch eine positive Wahlabsicht angibt. Auch sind Personen dieser Gruppe überdurchschnittlich häufig berufstätig, zwischen 40 und 49 Jahre alt und schätzen sowohl die eigene als auch die allgemeine wirtschaftliche Lage als gut ein. Da die jüngste Altersgruppe und diejenige zwischen 40 und 50 Jahren mit der Bildungsdimension korrelieren, lässt sich die zweite Achse lediglich über die Befragten mit 50 Jahren und älter im oberen und die Altersgruppe von 30 bis 39 Jahren im unteren Abschnitt beschreiben, so dass es sich hierbei nicht um eine ‚reine‘ Altersdimension handelt. Auf einer dritten, hier graphisch nicht abgebildeten Dimension, die weitere 12% der Gesamtvarianz erklärt, lädt die Gruppe der Home-Zone-Nutzer, die sich überdurchschnittlich häufig in der Ausbildung oder im Studium befindet, auf der einen Seite und die Gruppe der Nichterwerbstätigen auf der anderen Seite, der keine weitere Ausprägung zugeordnet werden kann.

Die unterschiedlichen Vorkommenshäufigkeiten der sozialstrukturellen und sozio-demographischen Variablen in den fünf Gruppen von Telefonnutzern unterscheiden sich in beiden Stichproben. Bei den im Telefonbuch eingetragenen Befragten scheint es hauptsächlich eine Frage des Alters und weniger der Bildung zu sein, welcher bzw. ob ein Festnetzanschluss im Haushalt vorliegt, während bei den Befragten mit generierten Mobilfunknummern diese mehr mit der Höhe des Bildungsabschlusses und den politischen Einstellungen differiert. Die exklusiven Mobilfunknutzer aus der Stichprobe mit generierten Nummern scheinen sich, so ein weiteres vorsichtiges Fazit, von denen mit im Telefonbuch eingetragenen Mobilfunknummern, bezogen auf den Bildungsabschluss und das Interesse an Politik sowie an politischer Partizipation, zu unterscheiden. Bei der Gruppe der exklusiven Mobilfunknummern mit Eintrag im Telefonbuch, die sich häufig in der Ausbildung oder im Studium befinden, könnte es sich bei der ‚Abkehr‘ vom Festnetz um einen Lebenszykluseffekt handeln.

Allerdings erweisen sich im Rahmen einer durchgeführten Logit-Analyse (Tabelle 6, Modell 1) weder der Erwerbsstatus noch die Einschätzung der eigenen wirtschaftlichen Lage als signifikante Prädiktoren der exklusiven Mobilfunknutzung unter den eingetragenen Mobilfunkteilnehmern.

Tabelle 6 Binäre Logit-Analyse für die exklusive Mobilfunknutzung^a

	Modell 1			Modell 2		
	Wahlberechtigte B	Wahlberechtigte Wald	Wahlberechtigte Sig.	Wahlberechtigte B	Wahlberechtigte Wald	Wahlberechtigte Sig.
Konstanter Term	-2,354	13,277	0,000	-1,113	3,647	0,056
Gebiet (<i>Ref. Ost</i>)	-0,718	9,604	0,002	-0,641	9,772	0,002
Ortsgröße						
bis 20.000 Einw.	0,501	2,740	0,098	-0,733	12,426	0,000
bis 100.000 Einw.	0,838	6,684	0,010	-0,237	1,023	0,312
<i>Ref. mehr als 100.000 Einw.</i>	0	.	.	0	.	.
Geschlecht (<i>Ref. Frau</i>)	-0,018	0,005	0,941	0,780	15,805	0,000
Alter						
18 bis 29 Jahre	1,135	11,174	0,001	1,397	16,007	0,000
30 bis 39 Jahre	0,292	0,702	0,402	0,923	6,848	0,009
40 bis 49 Jahre	-0,308	0,627	0,428	0,551	2,368	0,124
<i>Ref. 50 Jahre und älter</i>	0	.	.	0	.	.
Bildung						
ohne/Volks-/Hauptschulab.	0,605	3,873	0,049	0,754	7,907	0,005
mittlere Reife/POS	0,228	0,817	0,366	0,470	3,858	0,050
<i>Ref. Abitur</i>	0	.	.	0	.	.
Erwerbsstatus						
arbeitslos	0,634	1,538	0,215	1,012	11,155	0,001
Rentner	-0,090	0,034	0,853	0,702	2,354	0,125
in Ausbildung/Hochschule	-0,243	0,378	0,539	0,029	,009	0,926
nicht erwerbstätig/Haushalt	-0,638	1,031	0,310	-0,399	,784	0,376
<i>Ref. erwerbstätig</i>	0	.	.	0	.	.
Eigene wirtschaftl. Lage ^b	-0,063	0,168	0,682	-0,294	5,110	0,024
Politisches Interesse ^c	-0,125	1,537	0,215	-0,272	9,680	0,002
Fallzahl	N = 1.477			N = 1.268		
Pseudo R2	7,3%			11,9%		
Basis: designgewichtete Daten						

^a Abhängige Variable: 0 Mobilfunk+Festnetz (analog, ISDN, Homezone, VoIP, Kabel), 1 exkl. Mobilfunknutzung.

^b Eigene wirtschaftliche Lage: 1 schlecht, 2 teils/teils, 3 gut.

^c Politisches Interesse: 1 gar nicht, 2 kaum, 3 etwas, 4 stark, 5 sehr stark.

Dagegen bestätigen sich die Ergebnisse der Korrespondenzanalyse für die nicht eingetragenen exklusiven Mobilfunknutzer (Tabelle 6, Modell 2). Die Gruppe der nicht eingetragenen exklusiven Mobilfunknutzer weist relativ häufig Merkmale sozialer Benachteiligung auf, wie geringe Bildung und Arbeitslosigkeit sowie damit verbundenen politischen Pessimismus. Es handelt sich um Personen, die in größeren Städten wohnen, eher männlich sind und den jüngeren Altersgruppen angehören.¹¹

4 Kombiniertes Stichprobendesign von Festnetz- und Mobilfunkstichprobe

Die Festnetzstichprobe des Politbarometers vom Dezember 2006 und die Mobilfunkstichprobe mit den generierten Nummern aus dem FGW-Mobilfunkpool wurden nach dem Dual-Frame-Ansatz zusammengespielt.¹² Unsere Grundgesamtheit umfasst die wahlberechtigte Bevölkerung. Die Transformationsgewichte berechnen sich nach der Formel:

$$1/(k_i^F \cdot \frac{1}{z_i} + \frac{m^C}{M^C} \cdot \frac{M^F}{m^F} \cdot k_i^C)$$

mit:

M^F = Anzahl Festnetznummern im Auswahlrahmen

m^F = Festnetznummern in Stichprobe

k_i^F = Zahl Festnetznummern, über die der Haushalt erreichbar ist

z_i = Zahl der zur Grundgesamtheit gehörenden Haushaltsmitglieder

M^C = Anzahl Mobilfunknummern im Auswahlrahmen

m^C = Mobilfunknummern in Stichprobe

k_i^C = Zahl Mobilfunknummern, über die der Befragte erreicht werden kann

Neben der Problematik der Erfassung der telefonischen Erreichbarkeiten und deren Validität (vgl. Meier 2007) stellt sich die Frage nach der Zahl der Nummern in den Auswahlrahmen (M^F und M^C) und den Stichprobengrößen. Hoffmann verwendet in seinem Test ein Verhältnis von 55 Mio. Festnetznummern zu 80 Mio. Mobil-

11 Vergleichbare Ergebnisse konnten auch anhand der f2f-Umfrage ermittelt werden. Bei den exklusiven Mobilfunknutzern der f2f-Umfrage handelt es sich überdurchschnittlich häufig um junge Frauen und Männer zwischen 24 und 29 Jahren aus Ostdeutschland, die überproportional oft arbeitslos sind.

12 Das Gewichtungungsverfahren und seine Annahmen werden in Gabler/Ayhan (2007) dargestellt, eine Anwendung in Hoffmann (2007). Die Analyse von Hoffmann bezieht sich allerdings auf eine andere Grundgesamtheit: die Wohnbevölkerung im Alter zwischen 18 und 64 Jahren. Die Erhebungen fanden im ersten Halbjahr 2006 statt.

funknummern. Diese Zahlen entsprechen etwa der Anzahl der Telefonkanäle, welche die Bundesnetzagentur (2007: 59) für das Jahr 2006 ausweist.¹³

Unsere Grundgesamtheit umfasst rund 61,9 Mio. Wahlberechtigte (Stand: Bundestagswahl 2005). Die Daten unserer f2f-Umfrage zeigen, dass nur rund 75% dieser Personen Handynutzer sind und ihre durchschnittliche Erreichbarkeit bei 1,1 Mobilfunknummern pro Person liegt. Daraus resultiert bei rund 80 Mio. Nummern insgesamt ein Mobilfunkauswahlrahmen von 51 Mio. Handynummern. Den Verhältnissen der Wahlberechtigten in West und Ost entsprechend umfasst der Auswahlrahmen im Westen 41,4 Mio. Nummern, im Osten 9,6 Mio. Hinsichtlich des Festnetzauswahlrahmens wird der Anteil der Haushalte ohne Personen mit deutscher Staatsangehörigkeit im Westen auf etwa 8% und im Osten auf etwa 2% geschätzt (Quelle: Mikrozensus-SUF 2005, eigene Berechnung). Daraus resultiert bei 55 Mio. Festnetznummern insgesamt ein Auswahlrahmen für Haushalte mit Wahlberechtigten von rund 41 Mio. Festnetznummern im Westen (55 Mio. \cdot 0,812 \cdot 0,92) und von rund 10 Mio. im Osten (55 Mio. \cdot 0,188 \cdot 0,98). Folgende Transformationsgewichte (TR) werden somit in den nachfolgenden Analysen verwendet:

$$\text{TR}_F \text{ (Festnetzstichprobe)} \quad 1/\pi_i = 1/(k_i^F \cdot \frac{1}{z_i})$$

$$\text{TR}_C \text{ (Mobilfunkstichprobe)} \quad 1/\pi_i = 1/k_i^C$$

TR_DFW (Dual-Frame West)

$$1/\pi_i = 1/(k_i^F \cdot \frac{1}{z_i}) + \frac{41}{965} \cdot \frac{998}{41,4} \cdot k_i^C = 1/(k_i^F \cdot \frac{1}{z_i}) + 1,0242 \cdot k_i^C$$

TR_DFO (Dual-Frame Ost)

$$1/\pi_i = 1/(k_i^F \cdot \frac{1}{z_i}) + \frac{10}{668} \cdot \frac{270}{9,6} \cdot k_i^C = 1/(k_i^F \cdot \frac{1}{z_i}) + 0,4210 \cdot k_i^C$$

13 Allerdings beschränkt sich die Grundgesamtheit bei Hoffmann auf die Wohnbevölkerung von 18 bis 64 Jahren und umfasst damit nur einen Teil der in den Auswahlrahmen angenommenen Größenordnungen. Zum Beispiel enthält der Festnetzrahmen die Nummern von rund 8,6 Mio. Haushalten (22% der 39,2 Mio. Haushalte insgesamt im Jahr 2005), in denen ausschließlich Personen ab 65 Jahren wohnen (vgl. Statistisches Bundesamt 2006). Entsprechend dürfte auch der Auswahlrahmen für den Mobilfunkbereich überschätzt sein, da dieser sowohl die Nummern von Personen zwischen 14 und 18 Jahren (ca. 3,9 Mio. Personen) als auch derjenigen ab 65 Jahren (ca. 15,4 Mio. Personen) umfasst.

Tabelle 7 zeigt einige der Transformationsgewichte im Dual-Frame-Ansatz. Den Exklusivnutzern (Mobilonlys bzw. Nur-Festnetznutzer) werden dabei relativ hohe Transformationsgewichte zugewiesen.

Tabelle 7 Transformationsgewichte

Haushaltsgröße	Festnetznummern	Mobilnummern	TR_DFW	TR_DFO
2	0	2	0,49	1,19
1	0	1	0,98	2,38
2	1	0	2,0	2,0
4	1	0	4,0	4,0
1	1	1	0,49	0,70
2	3	1	0,40	0,52
4	1	1	0,78	1,49

Tabelle 8 gibt die Ergebnisse der entsprechenden Designgewichtungen wieder. Spalte 1 und 2 zeigen die Ergebnisse der Festnetzstichprobe und der Mobilfunkstichprobe, Spalte 3 die Ergebnisse des kombinierten Stichprobendesigns. Diese Daten wurden zusätzlich noch korrekturgewichtet (Spalte 4), d. h. hinsichtlich Alter x Geschlecht x Bildung an den Mikrozensus 2005 (Spalte 5a) angepasst.

Beim kombinierten Stichprobendesign ergänzen sich Festnetz- und Mobilfunkstichprobe nur zum Teil (Vergleich Spalte 3 und 5a): Designbedingt werden zum Beispiel die unter 30-jährigen Männer im Osten besser erreicht und ihr Anteil entspricht nun dem der amtlichen Statistik. Bei den über 50-Jährigen ist dagegen eine Verschlechterung hinsichtlich ihres Anteils zu beobachten. Entsprechend sinkt der Anteil der Rentner. Die höheren Bildungsgrade sinken anteilmäßig, aber die Bildungsverteilung nähert sich nicht in nennenswertem Umfang der amtlichen Statistik an.

Hinsichtlich der politisch-inhaltlichen Variablen ist zu beobachten (Vergleich Spalte 1 und 3), dass der Anteil der Personen, die politisch kaum oder gar nicht interessiert sind, durch die Kombination mit der Mobilfunkstichprobe in West und Ost designbedingt um etwa fünf Prozentpunkte ansteigt, die potenzielle Wahlbeteiligung sinkt entsprechend. Der Anteil der rechten und sonstigen Parteien steigt an. Diese Ergebnisse deuten darauf hin, dass über den Mobilfunkauswahlrahmen durchaus Befragte gewonnen werden, die in politischen Umfragen tendenziell unterrepräsentiert sind.

Tabelle 8 Gewichtungseffekte^a

	Festnetzstichprobe design- gewichtet TR_F		Mobilfunkstichprobe design- gewichtet TR_C		Dual-Frame design- gewichtet TR_DF		Dual-Frame korrektur- gewichtet (Alter, Bildung, Geschlecht)		Mikrozensus 2005 (scientific use file) Deutsche ab 18 Jahren 5a	
	1		2		3		4		5a	
	West	Ost	West	Ost	West	Ost	West	Ost	West	Ost
Fallzahl	965	668	998	270	1963	938	1963	938		
Alter x Geschlecht										
Männer –29	7,5	6,6	18,6	16,7	12,3	9,9	8,5	9,6	8,2	9,6
–49	23,8	16,5	26,8	23,0	24,0	17,4	18,3	17,6	18,3	17,6
50+	17,8	23,4	10,3	11,7	15,7	19,9	21,3	21,2	21,5	21,2
Frauen –29	9,4	6,5	13,9	17,6	10,7	10,4	8,4	8,6	7,9	8,2
–49	22,3	20,9	21,9	21,3	20,7	19,3	17,7	16,1	18,1	16,7
50+	19,2	26,1	8,4	9,6	16,6	23,1	25,9	26,9	26,0	26,6
Bildung										
kein Abschluss/noch Schüler	1,8	2,1	3,1	4,8	2,1	2,9	2,8	3,0	3,7	2,5
Volks- und Hauptschulabschluss	28,4	16,8	27,5	16,1	30,7	19,6	47,8	30,2	47,0	29,9
mittlere Reife / POS	31,9	48,0	35,5	53,6	33,2	48,4	24,5	46,6	24,4	46,5
(Fach-) Abitur	16,4	9,5	15,2	8,7	14,0	8,2	11,6	7,9	12,6	8,9
Fach-/ Hochschulabschluss	21,1	23,4	17,8	15,1	19,3	20,1	12,5	11,7	11,4	11,5
keine Angabe	0,4	0,3	1,0	1,9	0,8	0,8	2,7	0,7	0,9	0,6
Erwerbstätigkeit										
Arbeitslos	4,1	9,3	5,5	10,1	5,6	10,5	6,0	10,2	4,8	11,1
Rente	21,1	29,0	5,9	6,4	17,4	24,4	27,6	31,0	26,4	31,4
Ausbildung/Studium	7,6	5,3	11,1	10,7	8,3	7,0	5,1	5,5	2,4	2,8
nicht erwerbstätig/Erzieh. erwerbstätig (voll/teil/kurz)	9,8	4,5	8,0	5,6	9,2	4,3	10,3	4,3	12,2	4,6
	57,2	51,8	69,1	66,8	59,3	53,6	50,7	48,8	54,2	50,1
	Festnetzstichprobe design- gewichtet TR_F		Mobilfunkstichprobe design- gewichtet TR_C		Dual-Frame design- gewichtet TR_DF		Dual-Frame korrektur- gewichtet (Alter, Bildung, Geschlecht)		Festnetzstich- probe korrektur- gewichtet (Alter, Bildung, Geschlecht)	
	1		2		3		4		5b	
	West	Ost	West	Ost	West	Ost	West	Ost	West	Ost
Fallzahl	965	668	998	270	1963	938	1963	938	965	668
Politikinteresse										
(sehr) stark	40,3	42,4	36,4	29,7	37,1	36,5	35,0	35,3	37,7	38,0
etwas	42,8	37,6	39,2	41,2	40,6	38,4	40,7	37,3	43,4	37,4
kaum, gar nicht	17,0	20,0	24,4	29,1	22,0	24,8	23,9	27,2	19,0	24,6
Gehe zur Wahl										
ja	83,8	73,3	72,9	64,9	78,3	68,2	77,8	66,0	81,5	69,7
weiß nicht	2,7	5,5	6,3	5,4	4,5	5,2	4,4	5,9	2,8	6,2
nein	13,5	21,2	20,8	29,6	17,1	26,0	17,7	27,7	15,7	24,1
Wahlabsicht (gültige)										
CDU/CSU	41,3	31,6	30,6	39,0	38,0	34,7	40,6	34,5	42,7	31,9
SPD	29,2	30,1	37,7	23,5	31,8	28,8	33,9	30,1	31,4	30,6
FDP	9,9	4,4	11,2	7,4	9,9	5,2	7,6	4,9	8,8	4,6
Linke	4,1	23,7	2,3	13,7	3,3	19,7	3,4	20,9	4,0	24,5
Grüne	12,8	6,3	11,9	5,1	12,3	5,9	9,9	5,1	9,9	4,5
Rechte	1,2	1,0	2,4	5,4	1,8	2,1	1,5	1,8	1,3	0,9
Andere	1,5	2,9	3,9	5,9	2,9	3,6	3,0	2,7	2,0	3,0
Variationskoeffizient (Gewichte)	0,53	0,46	0,23	0,21	0,63	0,53	0,76	0,68	0,69	0,68

^a Fehlend zu 100% = keine Angabe

Im Vergleich der korrekturgewichteten Ergebnisse von Festnetzstichprobe und Dual-Frame-Ansatz (Spalte 4 und 5b) fallen die Unterschiede bezüglich der politisch-inhaltlichen Variablen allerdings etwas moderater aus. In der politischen Stimmung erhält die CDU/CSU im Westen zwei Prozentpunkte weniger, die FDP einen weniger, die SPD zweieinhalb mehr. Der größte Unterschied ist bei der Linken im Osten zu verzeichnen. Gegenüber den Festnetzstichprobenergebnissen werden knapp vier Prozentpunkte weniger berechnet

5 Zusammenfassung und Fazit

Die zunehmende Mobilfunknutzung kann für repräsentative Bevölkerungsumfragen, die auf Festnetzstichproben basieren, zukünftig zum Problem werden. Der Anteil an exklusiven Mobilfunknutzern, die systematisch aus Festnetzstichproben ausgeschlossen werden, scheint aber eher zu stagnieren und nicht weiter zuzunehmen. In einer bundesweiten repräsentativen f2f-Befragung wurde Ende des Jahres 2006 ein Anteil von 5,7% Personen in der Wohnbevölkerung ermittelt, die nicht mehr über einen Festnetzanschluss in ihrem Haushalt verfügen. Diese Situation stellt den Anspruch an Repräsentativität von Festnetzstichproben sicher nicht infrage.

Eingangs wurde die Frage gestellt, ob es sinnvoll erscheint, zur besseren Erreichbarkeit mobiler Personenkreise und vor allem zur prinzipiellen Erreichbarkeit exklusiver Mobilfunknutzer die rund 2 Mio. im Telefonbuch eingetragenen Mobilfunknutzer in Festnetzstichproben zu integrieren. Mobilfunkteilnehmer, die ihre Handynummer im Telefonbuch eingetragen haben, unterscheiden sich jedoch in sozialstrukturellen Merkmalen von denjenigen mit nicht eingetragenen Nummern. Die Handynutzer mit eingetragener Rufnummer sind höher gebildet und auch häufiger Vollzeit berufstätig. Diese Unterschiede in den Stichproben der ins Telefonbuch eingetragenen Handynutzer zu den Befragten mit generierten Handynummern brachten uns nun weiter zu der Betrachtung der exklusiven Mobilfunknutzer in beiden Stichproben, diejenige Gruppe also, die wir bei Telefonumfragen nicht erreichen können. Abgesehen von dem Aufwand, den man über Screeningverfahren betreiben müsste, um den entsprechenden Anteil an exklusiven Mobilfunknutzern zu erhalten, wären die im Telefonbuch eingetragenen einer regionalen Schicht zuzuordnen. Doch vor allem die exklusiven Mobilfunknutzer, welche im Telefonbuch eingetragen sind, unterscheiden sich von denjenigen, die dies nicht wünschen. Exklusive Mobilfunknutzer mit Eintrag im Telefonbuch sind eher jüngere Personen, die sich in der Ausbildung oder im Studium befinden, sich eher für Politik interessieren und ihre eigene wirtschaftliche Lage als weniger kritisch betrachten. Ihre Abkehr

vom Festnetz könnte sich nach diesem Lebensabschnitt wieder wandeln und könnte vorsichtig als Lebenszykluseffekt interpretiert werden, während die andere Gruppe exklusiver Mobilfunknutzer relativ häufig Merkmale sozialer Benachteiligung wie geringe Bildung und Arbeitslosigkeit und damit verbunden politischen Pessimismus aufweist, was ebenfalls anhand der repräsentativen f2f-Umfrage gestützt wurde. Eine Integration von im Telefonbuch eingetragenen Handynummern, um die regionale Schichtung einer Telefonstichprobe zu ermöglichen, scheint uns anhand dieser Ergebnisse als nicht sinnvoll. Die beiden Gruppen von Mobilfunknutzern unterscheiden sich anhand von Merkmalen, die mit den in sozialwissenschaftlichen Umfragen untersuchten Variablen bekanntermaßen korrelieren. Erst die weitere Entwicklung der exklusiven Mobilfunknutzung wird also zeigen, welche Konsequenzen sich für repräsentative Festnetzstichproben ergeben.

Durch den Dual-Frame-Ansatz werden offensichtlich einige unterrepräsentierte und gleichzeitig schwer zu befragende Gruppen (junge Männer im Osten, politisch eher wenig/gar nicht Interessierte, Wähler sonstiger/rechter Parteien) besser abgebildet. Andererseits werden bestimmte Gruppen designbedingt unterrepräsentiert (Renter, über 50-Jährige) und der grundsätzliche Bildungsbias wird auch durch die Mobilfunkstichprobe designbedingt nicht wirklich beseitigt. Ein Vergleich zwischen den korrekturgewichteten Daten der Festnetzstichprobe und des Dual-Frame-Ansatzes zeigt, dass die Veränderungen der politisch-inhaltlichen Variablen letztlich moderat ausfallen.

Weitere Forschungen sind notwendig im Hinblick auf den Einsatz eines wesentlich umfangreicheren Fragebogens im Mobilfunkbereich und bezüglich der Frage, inwieweit Terminvereinbarungen und -einhaltungen über das Handy möglich sind. Abgesehen von Kostenüberlegungen, dürfte es zum dauerhaften Einsatz von Mobilfunkstichproben erst kommen und auch inhaltlich sinnvoll sein, wenn der Anteil der Mobilonlys weitaus stärker zunimmt.

Literatur

- Blasius, J., 1988: Zur Stabilität von Ergebnissen bei der Korrespondenzanalyse. *ZA-Information* 23: 47-62.
- Blasius, J., 2001: *Korrespondenzanalyse*. München: Oldenbourg.
- Bundesnetzagentur, 2007: *Jahresbericht 2006*. Bonn.
- Gabler, S. und Ö. Ayhan, 2007: Gewichtung bei Erhebungen im Festnetz und über Mobilfunk: Ein Dual-Frame-Ansatz. S. 39-46 in: S. Gabler und S. Häder (Hg.): *Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung*. ZUMA-Nachrichten Spezial Band 13. Mannheim.
- Gabler, S. und S. Häder (Hg.), 2007: *Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung*. ZUMA-Nachrichten Spezial Band 13. Mannheim.

- Glemser, A., 2007: Mobilfunknutzung in Deutschland. Eine Herausforderung für die Stichprobenbildung in der Markt- und Sozialforschung. S. 7-24 in: S. Gabler und S. Häder (Hg.): Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung. ZUMA-Nachrichten Spezial Band 13. Mannheim.
- Häder, S. und A. Glemser, 2006: Stichprobenziehung für Telefonumfragen in Deutschland. S. 148-171 in: A. Diekmann (Hg.): Methoden der Sozialforschung. Kölner Zeitschrift für Soziologie und Sozialpsychologie Sonderband 44. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Heckel, C., 2007: Weiterentwicklung der ADM-CATI-Auswahlgrundlagen. S. 25-38 in: S. Gabler und S. Häder (Hg.): Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung. ZUMA-Nachrichten Spezial Band 13. Mannheim.
- Hoffmann, H., 2007: Kombinierte Stichproben für Telefonumfragen – Ansätze in Europa. S. 47-58 in: S. Gabler und S. Häder (Hg.): Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung. ZUMA-Nachrichten Spezial Band 13. Mannheim.
- Meier, G., 2007: Validierung eines Fragebogens zur Erfassung der Anzahl von Telefonnummern. S. 91-104 in: S. Gabler und S. Häder (Hg.): Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung. ZUMA-Nachrichten Spezial Band 13. Mannheim.
- Schneiderat, G. und S. Häder, 2007: Anlage und Vorstudien des DFG-Projektes „Telefonbefragungen in der Allgemeinbevölkerung über das Mobilfunknetz“. S. 59-80 in: S. Gabler und S. Häder (Hg.): Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung. ZUMA-Nachrichten Spezial Band 13. Mannheim.
- Schulte, W., 1997: Telefon- und face-to-face-Umfragen und ihre Stichproben. S. 148-195 in: S. Gabler und J. Hoffmeier-Zlotnik (Hg.): Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag GmbH.
- Schroth, Y., 2003: Die Stichprobe. S. 37-50 in: A. Wüst (Hg.): Politbarometer. Opladen: Leske & Budrich.
- Statistisches Bundesamt, 2003: Wirtschaftsrechnungen. Einkommens- und Verbrauchsstichprobe – Ausstattung privater Haushalte mit langlebigen Gebrauchsgütern. Fachserie 15, Heft 1. Wiesbaden.
- Statistisches Bundesamt, 2006: Zahl der Woche Nr. 39 vom 26.09.2006. Wiesbaden.

Korrespondenzadressen: Stefan Hunsicker &
Yvonne Schroth
Forschungsgruppe Wahlen e.V.
N7, 13-15
68161 Mannheim
stefan.hunsicker@forschungsgruppe.de
yvonne.schroth@forschungsgruppe.de

Die Messung von Nutzen und subjektiven Wahrscheinlichkeiten

Ein Vorschlag zur Operationalisierung der Rational Choice Theorie

The measurement of utility and subjective probability

A proposal for the operationalization of rational choice theory

Henning Best

Zusammenfassung

Bislang sind in der empirischen Sozialforschung nur recht wenige direkte Anwendungen der Theorie rationalen Handelns (RCT) zu finden. Um solche empirischen Anwendungen zu erleichtern, wird ein ausgearbeiteter Vorschlag zur Operationalisierung und Messung der zentralen RC-Variablen präsentiert. Im Zentrum steht die Messung von Nutzen und subjektiven Wahrscheinlichkeiten im Sinne der Wert-Erwartungs- / SEU-Theorie. Am Beispiel einer Untersuchung zur Umstellung auf ökologische Landwirtschaft werden die Ergebnisse der Messmethode diskutiert.

Abstract

To date only a small number of direct applications of rational choice theory (RCT) can be found in empirical social research. In order to facilitate such empirical applications an elaborate proposal for the measurement and operationalization of central RC-variables is presented. The paper focuses on the measurement of utility and subjective probabilities, which in turn enable a direct empirical assessments of SEU-values. The measurement methodology and its result are discussed using an empirical study on the adoption of organic farming as an example.

Obwohl die Rational Choice Theorie (RCT) in den letzten Jahrzehnten in der Soziologie weite Verbreitung gefunden hat, wird die empirische Anwendung und direkte Überprüfung der Theorie noch immer eher stiefmütterlich behandelt. Zwar wurde die verstärkte Anwendung von Rational-Choice-Modellen in der empirischen Forschung immer wieder von prominenter Seite gefordert (so etwa Friedrichs et al. 1993; Goldthorpe 1996; Opp 1979, 1990), doch auch im Jahre 2004 konnte Brüderl noch ein „gespannte[s] Verhältnis von RC-Modellen und Umfrageforschung“ diagnostizieren (Brüderl 2004: 163).

Dieses gespannte Verhältnis wird in der Literatur auf eine Reihe von Faktoren zurückgeführt. Ein Grund ist sicherlich der typisch ökonomische Modellplatonismus, gepaart mit einer allgemeinen Abneigung gegenüber Umfragedaten in Teilen der Ökonomie (grundlegend Friedman 1953). Wichtiger erscheint jedoch für die Sozialwissenschaften die theoretische Kontroverse, ob inter-individuelle Variationen in Präferenzen geeignet sind, um Verhalten zu erklären (gleiches gilt für Variationen zwischen Gruppen wie etwa gesellschaftlichen Klassen). Nur wenn Präferenzunterschiede entscheidungsrelevant sind, gibt es überhaupt eine Notwendigkeit, Nutzen empirisch zu erheben. Hier lassen sich zwei grundlegende Positionen unterscheiden.

Vertreter der ersten Position (z. B. Diekmann 1996: 94ff.) postulieren, individuelle Präferenzen müssten nicht ermittelt werden, da sie konstant und somit nicht zur Erklärung von Handlungsentscheidungen geeignet sind (d. h. alle Menschen teilen die gleichen Präferenzen).¹ Zur Erklärung von Verhalten reichten Kenntnisse über die (objektiven oder subjektiven) Restriktionen vollkommen aus. Weiterhin wird argumentiert (etwa in Braun/Franzen 1995), es sei aufgrund erheblicher methodischer Probleme kaum möglich, Nutzen und Präferenzen valide zu messen. Dies sei aber auch nicht notwendig, da nicht die RCT als solches einer empirischen Prüfung zugänglich sein müsste, sondern lediglich die aus ihr ableitbaren Hypothesen. Eine leichte Variation dieser Argumentation liefert Lindenberg (1996a, 1996b). Er fordert, Präferenzen mehr oder weniger direkt aus seiner Theorie der sozialen Produktionsfunktionen abzuleiten, was mithin auch zu einer weitgehenden Konstanz der Präferenzen führt. Im Rahmen beider Argumentationen ist lediglich eine indirekte Überprüfung der Theorie möglich: Man formuliert Annahmen darüber, wie Randbedingungen der Handlung eine Entscheidung beeinflussen sollten, und erhebt sowohl die Randbedingungen als auch das Ergebnis der Entscheidung empirisch.

1 Diese Argumentation ist meines Ermessens nur haltbar, wenn ein sehr restriktiver Begriff von Präferenz verwendet wird. Präferenz wird hier als eine Art ‚Oberziel‘ verstanden. Üblich ist in diesem Zusammenhang die Bezugnahme auf Adam Smiths ‚soziale Anerkennung und physisches Wohlbefinden‘ (vgl. Smith 1926: 360f.), oder die Beschränkung auf ‚harte Nutzenfaktoren‘ wie Zeit oder Geld.

Die alternative Position schließlich besteht darin, Präferenzen als variabel und damit auch als entscheidungsrelevant anzunehmen. Über den genauen Inhalt der Präferenzen wird in dieser Variante der RCT a priori keine Aussage gemacht (Opp 1999 bezeichnet dies als die ‚weite Variante‘ der RCT). Hieraus wiederum ergibt sich die Notwendigkeit, Nutzen und weitere Bestandteile der RCT empirisch zu erheben (siehe z. B. Friedrichs et al. 1993; Kelle/Lüdemann 1995; Opp 1990; Opp/Friedrichs 1996). Folgt man dieser Strategie, können z. T. fragwürdige Annahmen darüber vermieden werden, ob Präferenzen konstant oder variabel über Individuen sind und welche Präferenzen für eine Entscheidung relevant sind. Im Gegensatz zur ersten geschilderten Grundposition (dem Modellierungs-Paradigma) ist die Theorie rationalen Handelns bei Anwendung dieser Strategie zudem einem direkten Test zugänglich.

Die direkte Anwendung der RCT in empirischen Untersuchungen stellt sich damit als die methodologisch strikteste, methodisch aber auch anspruchvollste Möglichkeit der Integration von empirischer Sozialforschung und Rational Choice dar. Lässt man sich auf eine empirische Anwendung der RCT ein, stellen sich vielfältige Fragen nach Möglichkeiten der Operationalisierung, die in der Literatur bislang nur unzureichend diskutiert wurden. Die Bearbeitung dieser Fragen wurde bereits Ende der 1970er Jahre von Opp (1979) als ein Desiderat der struktur-individualistischen Forschung formuliert. Auch haben Friedrichs et al. (1993), später auch Kunz (1994) sowie Kelle/Lüdemann (1995) versucht, eine solche Diskussion zu beginnen. Es ist jedoch nicht gelungen, eine systematische, kumulative Forschung zur Operationalisierung der RCT zu etablieren.

Ziel dieses Beitrages ist es, die Diskussion um empirische Überprüfungen der RCT neu aufzunehmen. Nachdem grundlegende theoretische Probleme bereits in den Arbeiten von Friedrichs et al. (1993) und Opp (1990) behandelt wurden, ist der Zugang dieses Artikels ein sehr praxisbezogener. Anhand eines konkreten Beispiels – der Umstellung auf ökologische Landwirtschaft – werden Möglichkeiten und Probleme der direkten Anwendung von RC-Modellen dargelegt und ein ausgearbeiteter, empirisch geprüfter Vorschlag zur Operationalisierung, Messung und Skalierung vorgestellt. Hiermit soll ein Handwerkszeug zur Verfügung gestellt werden, um empirische Forschung zukünftig zu erleichtern.

Der Beitrag ist wie folgt aufgebaut: Zunächst werden Grundlagen der RCT und der Wert-Erwartungs-Theorie rezipiert und die Kritik an einer direkten Anwendung der Theorie rationalen Handelns diskutiert. Nach einer kurzen Vorstellung des Untersuchungsbeispiels – ökologische Landwirtschaft – werden die konkreten Probleme, die sich bei der Operationalisierung der RCT stellen, diskutiert und sodann die in der Studie gewählte Operationalisierung vorgestellt. Um eine Validierung zu

bieten (oder zumindest die face-validity zu überprüfen) und um zu zeigen, welche inhaltlichen Erkenntnisse das Ergebnis einer direkten Anwendung der RCT sein können, werden zum Abschluss einige Ergebnisse der Studie zur Umstellung auf ökologische Landwirtschaft vorgestellt.²

1 Die Theorie rationalen Handelns

Die metatheoretische Grundlage der RCT ist der methodologische Individualismus: Kollektive, gesellschaftliche Phänomene (sog. Makrophänomene) resultieren aus dem Zusammenspiel des Handelns individueller Akteure. Für eine Erklärung dieser Makrophänomene ist dementsprechend ein Rückgriff auf das Handeln von Individuen, also auf die Mikroebene, notwendig (siehe Coleman 1995: 1–29). An diesem Punkt setzt die RCT ein. Sie stellt eine Mikrotheorie zur Verfügung, die angibt, wie sich Menschen im Durchschnitt verhalten. Menschliches Handeln wird hierbei als Entscheidungshandeln begriffen: Wenn ein Akteur handelt, muss er sich zwischen verschiedenen Handlungsalternativen entscheiden. Und dies, so die Theorie rationalen Handelns, tut er auf eine Weise, die angebbaren Regeln folgt. Rationales Handeln kann entsprechend als Handeln in Übereinstimmung mit einer Entscheidungsregel angesehen werden (vgl. Diekmann 1996: 91f.). Weitere Grundelemente rationaler Handlungstheorien sind nach Opp (1999: 173) Präferenzen und Restriktionen. Restriktionen bedingen die Handlungsalternativen, unter denen ein Akteur auswählen kann. Präferenzen wiederum beeinflussen das Ziel, das ein Akteur mit seiner Handlung verfolgt. Kurz gesagt postuliert die Theorie rationalen Handelns, dass ein Akteur unter Berücksichtigung von Restriktionen aus verschiedenen Handlungsalternativen diejenige auswählt, die seine Präferenzen am besten befriedigt.

Wie man sieht, lassen diese Grundvoraussetzungen einen weiten Spielraum für Interpretationen und Erweiterungen der Theorie. Um die RCT angemessen operationalisieren zu können, sind somit Präzisierungen notwendig. Die wichtigste Präzisierung betrifft die Entscheidungsregel, die Akteure anwenden. Im Rahmen dieses Beitrages wird auf die weit verbreitete subjective-expected-utility (SEU)-Theorie

2 Die Erhebung der empirischen Daten, die dieser Arbeit zugrunde liegen, wurde von der Fritz Thyssen Stiftung finanziert. Der Autor bedankt sich herzlich für die Unterstützung.

zurückgegriffen (siehe auch Fishburn 1981).³ Die SEU-Theorie besagt, dass unter mehreren Handlungsalternativen diejenige gewählt wird, von welcher der Akteur den höchsten Nettonutzen erwartet. Der Nettonutzen NN einer Handlungsalternative j wiederum kann als Summe der (Teil)Erwartungsnutzen aller Konsequenzen i , die der Akteur als Folge einer Entscheidung perzipiert, berechnet werden. Der (Teil)Erwartungsnutzen einer einzelnen Handlungskonsequenz ergibt sich als Produkt aus der Wahrscheinlichkeit p_i , dass die Konsequenz eintritt, und dem Nutzen U_i dieser Konsequenz. Dieser Nutzen U_i wird in der Literatur teilweise als Präferenz bezeichnet, was aber zu einer gewissen Begriffsverwirrung führen kann, da die Begriffe ‚Präferenzen‘, ‚Ziele‘, ‚Motive‘ und ‚Nutzen(aspekte)‘ meist deckungsgleich verwendet werden. Beispielsweise schreiben Opp/Roehl (1990: 8): „Die Präferenzen von Individuen sind in der Nutzentheorie die *Bewertungen*, d. h. die Nutzen, der Handlungsfolgen. Damit ist das Ausmaß gemeint, in dem Handlungsfolgen gewünscht oder auch nicht gewünscht werden.“⁴ Dieser Begriff deckt sich jedoch nur teilweise mit dem ökonomischen Präferenzbegriff (einführend Eisenführ/Weber 1994), der üblicherweise nur auf den Vergleich mindestens zweier Alternativen angewendet werden kann (im Sinne einer Rangfolge der Alternativen). Im Folgenden wird daher versucht, den Begriff ‚Präferenzen‘ weitgehend zu vermeiden und auf den präziseren Nutzenbegriff zurückzugreifen.

Zusammengefasst ergibt sich der Nettonutzen einer Alternative als

$$NN_j = \sum_{i=1}^n p_i U_i$$

Entsprechend muss für eine empirische Überprüfung im Fragebogen für alle Alternativen und alle Konsequenzen abgefragt werden, wie hoch die Wahrscheinlichkeit ist, dass eine Konsequenz eintritt und mit welchen Nutzen diese verbunden ist.

- 3 Die SEU-Theorie bietet sich für die vorliegende Untersuchung besonders an, da sich die Landwirte hinsichtlich der Folgen der Umstellung / Nicht-Umstellung in einer Unsicherheitssituation befinden. Die SEU-Theorie bietet eine elegante und recht sparsame Möglichkeit, diese Unsicherheit zu modellieren. Zudem ist sie sehr flexibel und kann leicht für spezielle Anforderungen angepasst werden.
- 4 Auf diese Begrifflichkeit habe ich in anderen Veröffentlichungen (z. B. Best 2006) selbst zurückgegriffen; auch dort wird U meist als Präferenz bezeichnet.

Nach Friedrichs et al. (1993) sind insbesondere folgende Fragen hinsichtlich der Operationalisierung der RCT von Relevanz:

1. Welche bzw. wie viele Handlungsalternativen und -konsequenzen werden von den Akteuren *wahrgenommen*?
2. Wie können die Handlungsalternativen und -konsequenzen *erhoben*, wie die Nutzen und Wahrscheinlichkeiten *gemessen* werden?
3. Wie sollen die Nutzen bzw. Kosten einer Alternative und die Eintrittswahrscheinlichkeiten einer Konsequenz *skaliert* werden?
4. Sollten die Konsequenzen zu *Dimensionen* zusammengefasst werden?

Bevor versucht wird, diese Fragen mit einer hohen Praxisnähe zu beantworten, wird im folgenden Abschnitt zunächst kurz auf die Kritik an direkten Anwendungen der RCT eingegangen.

2 Zur Kritik an direkten Anwendungen der RCT

Wie bereits in der Einleitung kurz angesprochen, gibt es in der Soziologie und der empirischen Sozialforschung eine Kontroverse über die Frage, wie empirisch mit der RCT umgegangen werden sollte. Die grundsätzliche Notwendigkeit einer Verbindung von Empirie und RCT wird hierbei von beiden Seiten anerkannt und beispielsweise von Goldthorpe (1996) als ‚soziologische Allianz‘ bezeichnet. Während, wie oben dargelegt, Friedrichs und Opp für eine direkte Anwendung und Überprüfung der RCT in empirischen Untersuchungen plädieren, werden in Teilen des RC-Lagers massive Vorbehalte gegen diese Vorgehensweise vorgebracht und statt dessen auf eine indirekte Modellbildung verwiesen. Beispielsweise schreibt Diekmann (1996: 98), direkte Anwendungen könnten die ‚hohe Messlatte‘ für RC-Erklärungen nicht erreichen, vielmehr sei es „die deduktive Stärke, die die Strategie indirekter Modellbildung vor den direkten Anwendungen der Nutzentheorie auszeichnet“. Darüber hinaus seien, so Diekmann, direkte Anwendungen von mindestens vier weiteren Problemen betroffen (S. 96):

- Die Messung von RC-Variablen führe zu methodischen Problemen, insbesondere sei das Skalenniveau der Variablen nicht ausreichend.
- Handlungsrestriktionen würden nicht oder über die subjektiven Wahrscheinlichkeiten nur implizit erfasst.
- Die statistische Umsetzung sei problematisch, da von einem deterministischen zu einem probabilistischen Modell übergegangen würde.
- Ex-post-Erklärungen, vor allem von historischen Ereignissen, seien bestenfalls illustrativ und mitunter tautologisch.

Diese Kritikpunkte werden im Wesentlichen auch von Braun/Franzen (1995) vertreten, die sich jedoch auf Probleme der empirischen Anwendung konzentrieren. Lindenberg (1996a, 1996b) wiederum präzisiert das Deduktions-Argument und fordert, Brückenannahmen (d. h. Annahmen über Präferenzen) theoretisch abzuleiten statt empirisch zu messen.

Zwar hat dieser Aufsatz es sich nicht zur Aufgabe gestellt, die grundsätzliche Auseinandersetzung zwischen Modellbildung und direkter empirischer Anwendung wieder aufzunehmen (vgl. hierzu aber Opp/Friedrichs 1996; Opp 1990, 1999), vielmehr will er einen praktischen Beitrag dazu leisten, empirische Probleme besser als bislang lösen zu können. Dennoch sollen einige Anmerkungen zu der Diskussion vorgebracht werden. So kann das Argument der Messprobleme zwar lang und breit diskutiert werden, es lässt sich aber nicht ohne praktische empirische Forschung lösen. Jede Messung ist zunächst einmal problematisch. Methodische Forschung hat daher das Ziel, diese Probleme einerseits zu erkennen, andererseits zu lösen. Gerade die Messung von RC-Variablen ist Gegenstand dieses Beitrages, wobei besonders das Skalenniveau der Messung, eines der wichtigsten messtechnischen Probleme, im weiteren Verlauf ausführlich behandelt wird. Auch das Argument, man prüfe in direkten empirischen Anwendungen ein probabilistisches RC-Modell statt eines deterministischen, ist nach Auffassung des Autors zwar im Grundsatz gerechtfertigt, spricht aber nicht per se gegen direkte empirische Anwendungen. Vielmehr sollte in solchen Anwendungen einerseits immer auch das deterministische Modell mit überprüft werden, und andererseits im Hinterkopf behalten werden, dass Überprüfungen einer probabilistischen Variante (z. B. in logistischen Regressionen) eben nicht den strengsten aller möglichen Tests darstellen. Das sicherlich schwerwiegendste der Argumente schließlich ist das wissenschaftstheoretische: Die vergleichsweise geringe deduktive Stärke der weiten Fassung der RCT. Opp (1999) diskutiert diesen Vorwurf ausführlich. Der Kern seines Argumentes ist, dass deduktive Erklärungen mit der engen Variante von RC, etwa dem klassischen homo oeconomicus-Modell, zwar mathematisch elegant und einfach seien, aber schlechterdings oft empirisch nicht zutreffend (etwa im Falle des Wahlparadoxes). Dies sei insbesondere deshalb nicht der Fall, wie die Forschung zu Anomalien zeige, da eine Reihe der restriktiven Annahmen der engen Variante von RC zurückgewiesen werden müssten. Unabhängig von dieser Gegen-Kritik bleibt festzuhalten, dass direkte Anwendungen der RCT eben nicht als konträr, sondern eher als komplementär zu modellierungs-orientierten Arbeiten stehen sollten. So können entsprechende empirische Untersuchungen gerade dazu beitragen, Anomalien der RCT besser zu verstehen und mitunter aufzulösen.

Da sich das Operationalisierungsbeispiel dieses Beitrages auf die Entscheidung über eine Umstellung auf ökologische Landwirtschaft bezieht, wird im folgenden Abschnitt zunächst kurz erläutert, was unter dieser Art der Landbewirtschaftung zu verstehen ist. Der Abschnitt liefert Hintergrundinformationen, welche die Details der Operationalisierung besser nachvollziehbar machen sollen.

3 Was ist ökologische Landwirtschaft?

Die ökologische Landwirtschaft ist eine besonders umweltfreundliche Form der Produktion von landwirtschaftlichen Gütern (siehe Mäder et al. 2002). Ziel der ökologischen Landwirtschaft ist es, den landwirtschaftlichen Betrieb als einen möglichst geschlossenen Kreislauf zu betreiben. Hierdurch soll es ermöglicht werden, die Bodenfruchtbarkeit langfristig zu erhalten, Umweltbelastungen zu vermeiden, Nutztiere artgerecht zu halten und Lebensmittel mit hohem gesundheitlichem Wert zu erzeugen. Zudem soll die ökologische Landwirtschaft einen Beitrag zur Lösung der globalen Energie- und Ressourcenprobleme leisten und die Grundlage für die Erhaltung bäuerlicher Strukturen in der Landwirtschaft legen (Bioland 2002). Nach Lampkin ist ökologische Landwirtschaft „best thought of as referring not to the type of inputs used, but to the concept of the farm as an organism, in which all the component parts – the soil minerals, organic matter, microorganisms, insects, plants, animals and humans – interact to create a coherent whole“ (Lampkin 1994: 5). Der wesentliche Unterschied zwischen ökologischer Landwirtschaft und anderen Formen der umweltfreundlichen, nachhaltigen Landbewirtschaftung ist diese holistische Sichtweise und die Existenz eines Regelwerkes, das es erlaubt, ökologische Landwirtschaft klar zu definieren.

Traditionell wurde und wird dieses Regelwerk selbstorganisiert von nicht-staatlichen Organisationen, so genannten Anbauverbänden, festgelegt (vgl. Bioland 2002; Demeter 2002; Naturland 2002). Die wichtigsten, international anerkannten Richtlinien stammen von der „International Federation of Organic Agriculture Movements“ (IFOAM 2002). Sie bestehen aus einer Vielzahl von Regeln zur Bodenbearbeitung, Schädlingsbekämpfung und Tierhaltung, die hier nicht detailliert wiedergegeben werden können. Ein wesentliches Merkmal der ökologischen Landwirtschaft ist jedoch, dass grundsätzlich auf den Einsatz gentechnisch veränderter Organismen, synthetischer Pestizide und Herbizide sowie auf synthetische Düngemittel verzichtet wird. Die Regeln zur umweltfreundlichen Bodenbewirtschaftung werden durch Richtlinien zur artgerechten Tierhaltung ergänzt.

Zusätzlich zu den in Selbstorganisation erarbeiteten Richtlinien trat mit der EU-Verordnung 2092/91 zum Ökolandbau im Jahr 1993 ein staatliches Regelwerk in Kraft, das per Gesetz festlegt, welche Anbauweise als ‚ökologische Landwirtschaft‘ bezeichnet werden darf und welche nicht (EC 1991). Diese staatlichen Richtlinien stellen im Allgemeinen weniger strenge Regeln auf als die anerkannten Anbauverbände dies tun.

Eine Bewirtschaftung des Betriebes nach den Regeln der ökologischen Landwirtschaft ermöglicht es dem Landwirt nicht nur, die Umwelt zu schonen, sondern gibt ihm vor allem auch die Möglichkeit, seine Produkte unter dem Prädikat ‚aus kontrolliert ökologischem Anbau‘ zu vermarkten. Produkte aus ökologischer Landwirtschaft erzielen auf dem Markt zum Teil deutlich höhere Preise als Produkte aus konventioneller Landwirtschaft.

4 Messung und Skalierung der RC-Variablen

4.1 Erhebung der Handlungsalternativen

Bezüglich seiner Handlungsalternativen steht der Landwirt nur scheinbar vor einer Dichotomie ‚Umstellung auf Ökolandbau oder weiter wie bisher‘. Denn es ist keineswegs sicher, dass alle Betriebsleiter überhaupt vor irgendeiner größeren Entscheidung über ihre Wirtschaftsweise stehen; und selbst wenn sie vor einer solchen Entscheidung stehen, muss der Ökolandbau für sie nicht unbedingt eine Handlungsmöglichkeit darstellen. Auch wenn ein Betriebsleiter die Umstellung auf Ökolandbau erwägt, ist die Entscheidungssituation nicht notwendig dichotom – es lassen sich ad hoc zahlreiche weitere Handlungsmöglichkeiten formulieren, z. B. die Aufgabe des Betriebes, Umstellung von Haupt- auf Nebenerwerbslandwirtschaft, die Erweiterung bzw. komplette Änderung der Produktpalette oder die Bereitstellung von Freizeitdienstleistungen wie ‚Urlaub auf dem Bauernhof‘. All diese und eine Vielzahl weiterer Alternativen kann ein Landwirt in Erwägung ziehen, wenn er entscheiden muss, wie sich sein Betrieb verändern soll.

Zwar wäre es denkbar, in einer qualitativen Vorstudie zu ermitteln, welche Handlungsmöglichkeiten von einer Mehrheit der Landwirte wahrgenommen werden, die dann in der Hauptuntersuchung verwendet werden könnten. Aus erhebungstechnischen Gründen kann – insbesondere in einer postalischen Befragung – den Befragten jedoch nur eine geringe Zahl von Alternativen zur Bewertung vorgelegt werden. Die Verwendung von drei Wahlmöglichkeiten ist nach Ansicht des Autors das Maximum,

das in postalischen Befragungen realisiert werden kann.⁵ Zudem kann angenommen werden, dass alle Alternativen zunächst mit dem Status Quo verglichen werden, sodass der Nutzen einer ‚weiter wie bisher‘-Alternative immer erhoben werden muss. Dies bedeutet jedoch, dass neben den Möglichkeiten ‚Ökolandbau‘ und ‚weiter wie bisher‘ nur eine zusätzliche Alternative in den Fragebogen aufgenommen werden könnte – das eingangs geschilderte Problem wäre damit nur in geringem Maße entschärft.

Abbildung 1 Abfrage der dritten Handlungsalternative

<p>Abgesehen davon, Ihren Betrieb so weiterzuführen wie bisher oder ihn auf ökologische Landwirtschaft umzustellen, sind noch weitere Entwicklungsmöglichkeiten denkbar. Haben Sie in letzter Zeit einmal über eine weitere Alternative nachgedacht?</p> <p>Falls ja, nennen Sie bitte die Alternative, die Sie alles in allem am attraktivsten fanden.</p> <p>¹ <input type="checkbox"/> Ja, und zwar _____</p> <p>-----</p> <p>² <input type="checkbox"/> Nein, keine weitere Alternative → <i>weiter zu Frage 36</i></p>	
35.	<p>Wenn Sie jetzt überlegen, wie Sie diese Alternative im Vergleich zu Ihrer bisherigen Betriebsform und dem ökologischen Landbau bewerten:</p> <p>A Finden Sie die oben genannte Möglichkeit alles in allem attraktiver, in etwa gleich oder weniger attraktiv als Ihren Betrieb so weiterzuführen wie bisher?</p> <p>¹ <input type="checkbox"/> attraktiver ² <input type="checkbox"/> etwa gleich ³ <input type="checkbox"/> weniger attraktiv</p> <p>B Und wie bewerten Sie die oben genannte Möglichkeit im Vergleich zum ökologischen Landbau?</p> <p>¹ <input type="checkbox"/> attraktiver ² <input type="checkbox"/> etwa gleich ³ <input type="checkbox"/> weniger attraktiv</p>

Eine andere Möglichkeit besteht darin, in der Befragung gar keine Alternativen vorzugeben, sondern diese offen abzufragen. Im Anschluss an die offene Frage würde der Befragte gebeten, die zwei oder drei seiner Meinung nach wichtigsten Alternativen zu bewerten. Da die Beantwortung von RC-Fragebatterien – wie sich im Folgenden zeigen wird – ohnehin recht aufwendig ist, kann davon ausgegangen werden, dass auch dieser Weg in Befragungen nur bedingt handhabbar ist, in postalischen Befragungen sicherlich überhaupt nicht.

5 Die multikriterielle Evaluation der Alternativen beansprucht einerseits viel Fragebogenplatz (im vorliegenden Beispiel eine A4-Seite pro Alternative) und ist andererseits für den Respondenten sehr aufwendig.

Es wurde daher ein Mittelweg zwischen der Vorgabe von Handlungsmöglichkeiten und einer offenen Abfrage gewählt: Die Alternativen ‚Ökolandbau‘ und ‚weiter wie bisher‘ wurden allen Teilnehmern vorgegeben. In zwei vorgeschalteten Filterfragen wurde abgefragt, ob der Befragte schon einmal über grundsätzliche Änderungen nachgedacht hat und, wenn ja, ob er eine Umstellung auf ökologische Landwirtschaft erwogen hat. Hierdurch kann die Untersuchung von individuell nicht relevanten Alternativen vermieden werden. Zusätzlich wurde im Anschluss an die beiden vorgegebenen Handlungsmöglichkeiten gefragt, ob die Befragten noch über eine weitere Alternative nachgedacht haben (siehe Abbildung 1). Hierdurch können Verfälschungen des Untersuchungsergebnisses durch eine verdeckte dritte Alternative vermieden und der Erhebungsaufwand in Grenzen gehalten werden.

4.2 Erhebung der Handlungskonsequenzen

Die Probleme, die sich bei der Erhebung von Handlungskonsequenzen ergeben, unterscheiden sich nur geringfügig von den oben geschilderten Problemen der Erhebung von Handlungsalternativen: Es ist möglich, dass unterschiedliche Akteure eine unterschiedliche Zahl von unterschiedlichen Konsequenzen wahrnehmen. Lüdemann (1992) schlägt daher vor, die Konsequenzen offen zu erheben, auch Bouffard (2002) verweist auf die Vorteile einer offenen Frage. Auch wenn die Vorteile eines solchen offenen Erhebungsverfahrens durchaus auf der Hand liegen, sprechen zwei gewichtige Argumente gegen dieses Vorgehen: Zum einen ist die offene Erhebung von Konsequenzen – gerade auch in postalischen Befragungen – nur mit großen Einschränkungen durchführbar. Zum anderen zeigen Experimente von Kahnemann/Tversky (1984; siehe auch Tversky/Kahnemann 1988), dass die (positive oder negative) Formulierung der Konsequenzen einen Bezugsrahmen aktiviert, der das Entscheidungsverhalten (und damit auch das Antwortverhalten) beeinflusst. Um einen eindeutigen Bezugsrahmen zu gewährleisten und Probleme bei der Durchführung der Befragung zu vermeiden, erscheint es also empfehlenswert, die Konsequenzen im Fragebogen vorzugeben.

Damit stellt sich jedoch die Frage, welche Konsequenzen vorgegeben werden sollen. Um die Handlungskonsequenzen festzustellen, schlagen Friedrichs et al. (1993) in Anlehnung an Ajzen/Fishbein (1980) vor, die (modal salienten) Konsequenzen in einem Pretest mit offenen Fragen zu ermitteln. In seinen Studien zu Protestverhalten hat Opp (siehe z. B. Opp et al. 1984: 37ff.) die Handlungskonsequenzen vornehmlich in Gruppeninterviews erhoben. Eine Zusammenfassung der unterschiedlichen Strategien liefern Kelle/Lüdemann (1995). Für die vorliegende Untersuchung wurde eine kombinierte Vorgehensweise gewählt: In einem schrift-

lichen Pretest wurden Handlungskonsequenzen offen abgefragt und die so ermittelten Konsequenzen um Ergebnisse einer Gruppendiskussion von Ökolandwirten und mehrerer qualitativer Einzelinterviews von Landwirten und Verbandsvertretern ergänzt.

4.3 Erhebung der Nutzen und der Wahrscheinlichkeiten

Insgesamt konnten in den Vorstudien elf modal saliente Konsequenzen ermittelt werden. Für jede Konsequenz wurde zunächst auf einer 5-stufigen Ratingskala abgefragt, wie hoch der Nutzen einer Konsequenz eingeschätzt wird (siehe Abbildung 2).

Abbildung 2 Abfrage der Nutzen

26. Jede Entscheidung über die Produktionsweise eines Betriebes bringt gewisse Vor- und Nachteile mit sich. Sie sehen hier eine Liste mit einigen dieser Vor- und Nachteile, die Folge einer solchen Entscheidung sein können.

Wie ist das bei Ihnen: Finden Sie die in der Liste genannten Punkte sehr gut, eher gut, teils/teils, eher schlecht oder sehr schlecht? Bitte kreuzen Sie für ihre Bewertung für jeden der Punkte an.

	sehr gut	eher gut	teils/teils	eher schlecht	sehr schlecht
Einfache und effektive Bekämpfung von Unkraut und Schädlingen	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Gute Preise für die Produkte	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Hoher Ertrag an landwirtschaftlichen Produkten	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
„Papierkram“ erledigen müssen	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Gesicherter Absatz der Produkte	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Abhängigkeit von Subventionen	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Sicherheit vor Lebensmittelskandalen	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Umweltfreundliche Produktionsweise	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Gutes Image als Landwirt in der Bevölkerung	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Ausreichend Freizeit	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Hohe Prämien / Zuschüsse	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Keine chemischen Spritzmittel verwenden	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Umbauten an den Stallungen vornehmen müssen	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Langfristige Sicherung des Fortbestehens des Betriebes	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

Im Anschluss sollten die subjektiven Wahrscheinlichkeiten, dass die Konsequenzen eintreten, getrennt für die Alternativen ‚weiter wie bisher‘ und ‚Umstellung auf Ökolandbau‘ auf einer ebenfalls 5-stufigen Ratingskala angegeben werden. Die Fragen mussten für Ökobauern und konventionelle Landwirte leicht unterschiedlich gestellt werden, um der Situation gerecht zu werden – schließlich ist eine Weiterführung des Status Quo zum Befragungszeitpunkt für Ökolandwirte etwas anderes als für konventionelle Landwirte. Beispielhaft ist in Abbildung 3 die Frage dargestellt, in der konventionelle Landwirte angeben sollten, wie hoch die subjektive Wahrscheinlichkeit der Handlungskonsequenzen für den Fall ist, dass der Betrieb wie zuvor weitergeführt wird.

Abbildung 3 Abfrage der subjektiven Wahrscheinlichkeiten (konventionelle Landwirte, Handlungsalternative ‚weiter wie bisher‘)

28. Meist hat man ja auch eine Vorstellung darüber, ob eine Folge eher eintreffen oder auch eher nicht eintreffen wird.

Stellen Sie sich vor, dass Sie Ihren Betrieb im Großen und Ganzen so weiterführen, wie Sie das jetzt tun. Für wie wahrscheinlich halten Sie es, dass die folgenden Dinge eintreffen?

Bitte lesen Sie sich die Liste durch und kreuzen Sie für jeden Punkt an, ob Sie finden, dass er sicher, recht wahrscheinlich, vielleicht, wenig wahrscheinlich oder keinesfalls eintreffen wird, wenn Sie Ihren Betrieb im Wesentlichen so weiterführen wie bisher.

	sicher	recht wahrschein- lich	vielleicht	wenig wahrschein- lich	keinesfalls
Einfache und effektive Bekämpfung von Unkraut und Schädlingen	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Gute Preise für die Produkte	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

4.4 Skalenniveau der Variablen und Produktbildung

Die Entscheidung über eine Handlungsalternative wird, so die SEU-Theorie, nicht getrennt von Nutzen und Wahrscheinlichkeiten beeinflusst, sondern von dem Produkt der beiden Variablen (genauer der Produktsomme). Da bei der Verwendung von Produktvariablen in statistischen Auswertungen das Skalenniveau der zugrundeliegenden Variablen von erheblicher Bedeutung ist, soll hierauf zunächst näher eingegangen werden.

Die sicherlich gängigsten statistischen Methoden in soziologischen Anwendungen sind verschiedene Varianten der (bivariaten) Korrelation und der (multi-variaten) Regression. Werden die Variablen des Produktterms auf Intervallskalenniveau gemessen, sind Korrelationen und Regressionskoeffizienten nicht invariant gegenüber Skalentransformationen (insbesondere hinsichtlich einer Verschiebung des Nullpunktes) und damit *inhaltlich bedeutungslos*. In Regressionen kann dieses Problem behoben werden, indem man den Produktterm als Interaktionseffekt auffasst und die Haupteffekte mit in die Regressionsgleichung aufnimmt (siehe z. B. Allison 1977; Evans 1991; Kunz 1994). Werden die Variablen auf Ratioskalenniveau gemessen, sind additive Transformationen (also eine Verschiebung des Nullpunktes) nicht zulässig. Bei Messung mit Ratioskalen tritt das oben geschilderte Problem somit nicht auf, entsprechend können ratioskalierte Produktterme in korrelations- und regressionsanalytischen Verfahren ohne weiteres verwendet werden. Es ist dann nicht unbedingt notwendig, die Haupteffekte in das Regressionsmodell einzuschließen. Da die SEU-Theorie mit Produkten von Variablen operiert (nämlich Nutzen mal Wahrscheinlichkeit), sind empirische Anwendungen von der geschilderten Beschränkung betroffen, wenn keine Messung auf entsprechend hohem Skalenniveau erfolgt.

Wie im vorherigen Abschnitt beschrieben, erfolgte die Messung der Nutzen und der Wahrscheinlichkeiten anhand von Ratingskalen. Es ist in der Literatur umstritten, zu welchem Skalenniveau eine solche Messung führt (siehe z. B. Mayntz et al. 1969: 58; Sarle 1995; Scheuch/Zehnpfennig 1974: 114). Üblicherweise wird diskutiert, ob eine Messung auf Ordinal- oder auf Intervallskalenniveau erfolgt, wobei die Annahme von Intervallskalenniveau meist als unproblematisch angesehen wird. Das Vorliegen von Ratioskalenniveau wird in der Regel nicht diskutiert, da für die meisten Anwendungen Intervallskalen ausreichend sind. Um gleiche Abstände zu den Kategorien zu gewährleisten, wurden die visuellen Abstände zwischen den Antwortkategorien im Fragebogen konstant gehalten. Hierdurch sollte den Befragten signalisiert werden, dass zwischen den Kategorien auch inhaltlich gleiche Abstände bestehen. Bei der Formulierung der Kategorienbenennungen wurde, soweit möglich, auf Skalierungsuntersuchungen von Rohrman (1978) zurückgegriffen und diejenigen Formulierungen gewählt, die sprachlich möglichst gleiche Abstände aufweisen. Insofern kann auch die Annahme einer Messung der Nutzen und Wahrscheinlichkeiten auf mindestens Intervallskalenniveau als gerechtfertigt angesehen werden.

Um Ratioskalenniveau annehmen zu können, muss zusätzlich ein natürlicher Nullpunkt vorliegen: Dies ist im Falle der Wahrscheinlichkeiten gegeben (vgl. Diekmann 2002: 255). Axiomatisch liegen Wahrscheinlichkeiten zwischen Null und Eins,

wobei ‚keinesfalls‘ mit einer Wahrscheinlichkeit von Null gleichzusetzen ist. Auch im Falle des Nutzens ist die Annahme eines festen Nullpunktes nach der Auffassung des Autors gerechtfertigt: Als Antwortskala waren u. a. ‚sehr schlecht‘, ‚schlecht‘, ‚gut‘ und ‚sehr gut‘ vorgegeben. ‚Schlecht‘ ist negativ und kann als Kosten interpretiert werden, ‚gut‘ dagegen ist positiv und entsprechend Nutzen. Und zwischen negativ und positiv, zwischen Kosten und Nutzen, liegt notwendigerweise ein Nullpunkt. Entsprechend musste der Kategorie ‚teils/teils‘ eine Null zugewiesen werden, die ‚schlechten‘ Kategorien mussten negative und die ‚guten‘ Kategorien positive Werte erhalten. Eine sog. unipolare Skalierung der Nutzenvariablen, wie sie in der Literatur teilweise diskutiert wird (vgl. Ajzen 1991: 192f.; Pratkanis 1989, die allerdings beide für eine bipolare Skalierung votieren), ist in diesem Sinne unangemessen (gleiches gilt für eine bipolare Skalierung der Wahrscheinlichkeiten). Entsprechend wurden die Nutzen zu einer bipolaren Skala von $-2 =$ ‚sehr schlecht‘ bis $+2 =$ ‚sehr gut‘, die Wahrscheinlichkeiten zu einer unipolaren Skala von $0-1$ (‚keinesfalls‘ bis ‚sicher‘) umcodiert.⁶

Da man für beide Skalen gleiche Abstände annehmen und einen festen Nullpunkt angeben kann, ist es nach Ansicht des Autors gerechtfertigt, von einer Messung auf Ratioskalenniveau auszugehen.

Damit ist es möglich, den Erwartungsnutzen-Beitrag der einzelnen Konsequenzen für beide Alternativen als Produkt von Nutzen und Eintrittswahrscheinlichkeit der Konsequenz zu berechnen. Dieser Teil-Erwartungsnutzen einer Konsequenz kann zwischen -2 und $+2$ variieren.

Die Unterschiedlichkeit der beiden Alternativen hinsichtlich einer Konsequenz wird im Folgenden als ND_i bezeichnet. Sie ist die Differenz der beiden Teil-Erwartungsnutzen einer Konsequenz. Auch ND_i kann zwischen -2 und $+2$ variieren, wobei ein positives Vorzeichen auf einen Vorteil der ökologischen Wirtschaftsweise, ein negatives Vorzeichen auf einen Vorteil der konventionellen Landwirtschaft hindeutet.

6 Die Nutzen können also die Werte -2 ; -1 ; 0 ; 1 ; und 2 ; die Wahrscheinlichkeiten die Werte 0 ; $0,25$; $0,5$; $0,75$ und 1 annehmen.

4.5 Offene Fragen: Summenbildung und Gewichtung der Konsequenzen

Alle bisherigen Überlegungen bezogen sich ausschließlich auf einzelne Konsequenzen der Entscheidung. Der Akteur entscheidet sich jedoch, so die SEU-Theorie, nicht aufgrund der Bewertung einzelner Konsequenzen, sondern im Hinblick auf eine Gesamtbewertung der Alternativen. Mathematisch wird dies über eine Summenbildung modelliert. Eine solche Summierung ist dann absolut unproblematisch, wenn die Einzelnutzen auf identischen Skalen gemessen werden (z. B. in einer Währung wie Euro).

Nun ist es aber nicht selbstverständlich, dass die Nutzen sehr unterschiedlicher Konsequenzen auf identischen Skalen gemessen werden. Mit anderen Worten: Wenn ein Landwirt angibt, er fände es ‚sehr gut‘, keine chemischen Spritzmittel benutzen zu müssen, ist das dann das gleiche wie wenn er es ‚sehr gut‘ findet, seinen Betrieb langfristig ökonomisch sichern zu können? Nach Kenntnis des Autors ist diese Frage in der Literatur bislang nicht oder nur unzureichend diskutiert worden. Andererseits wird in der Einstellungsforschung, die mit den gleichen methodischen Problemen konfrontiert ist, traditionell davon ausgegangen, dass Items, die auf der selben Antwortskala beantwortet werden sollen, von den Befragten in der Tat auf der *gleichen evaluativen Dimension* angeordnet werden (z. B. Fishbein/Ajzen 1975: 80). *In dieser Studie wird davon ausgegangen, dass diese Annahme auch für die Messung von Nutzen zutrifft.* Die Ergebnisse dieser Studie, um darauf vorzugreifen, deuten darauf hin, dass eine Operationalisierung auf Basis dieser Annahme zu sinnvollen Ergebnissen führt: So bewerten konventionelle Landwirte (also der tendenziell repräsentativere Teil der Stichprobe) etwa die langfristige Sicherung ihres Betriebes im Durchschnitt mit +1,17, den Verzicht auf chemische Spritzmittel hingegen nur mit -0,12. Insofern ist zwischen diesen beiden Konsequenzen durchaus eine Abstufung, nicht nur in der Richtung, sondern auch in der Stärke des mittleren Nutzens zu erkennen. Allerdings muss auch beachtet werden, dass eine bipolare, 5-stufige Antwortmöglichkeit nur sehr begrenzt Raum für Abstufungen lässt. Hier wäre es unter Umständen empfehlenswert, mehr Antwortkategorien zu wählen oder mit alternativen Messmethoden zu experimentieren.

Eine andere mögliche Herangehensweise ist es, die Gleichheitsannahme nicht zu treffen und davon auszugehen, die Nutzen würden von den Befragten in unterschiedlichen Einheiten gemessen. Als Konsequenz müssten die einzelnen Nutzen (bzw. die Produkte) vor der Summenbildung reskaliert werden.

In der vorliegenden Untersuchung wurde bewusst auf eine Reskalierung verzichtet, da die Messergebnisse – wie oben kurz skizziert – Hinweise auf gleiche

Skalen geben, und in Zweifelsfällen die mathematisch einfachere Variante bis auf weiteres die bessere Wahl ist. Es ist eine Aufgabe für die weitere Forschung zur Operationalisierung von RC, ob eine Reskalierung notwendig und sinnvoll ist – und wenn ja, nach welchen Kriterien diese erfolgen sollte. Prinzipiell wäre dabei zu beachten, dass die Reskalierung auf die Bildung einer gewichteten Summe der Produkte $p_i U_i$ beschränkt werden muss, da der Nullpunkt und die relativen Abstände der Antwortkategorien des Nutzens U_i nicht verändert werden sollten. Eine Optimierung über die Interkorrelationen der Items, wie sie etwa in einer kategorialen Hauptkomponentenanalyse (CatPCA) erfolgt, ist von daher nicht angebracht, da die skizzierte Transformation sich nicht auf die Korrelationsmatrix der einzelnen Teil-Erwartungsnutzen auswirkt. Denkbar wäre es hingegen, Gewichte für die Summenbildung zu suchen, die die Korrelation des SEU-Wertes mit der Entscheidung maximieren. Ein Startpunkt hierfür könnten Korrelationen (unter Umständen auch Regressionskoeffizienten) der Nutzen U_i oder der Produkte $p_i U_i$ mit der Entscheidung sein. Eine solche Gewichtung wäre aber mit Problemen statistischer und theoretischer Natur verbunden, die eine gesonderte Untersuchung erfordern. In jedem Fall ist es sinnvoll, zunächst zu prüfen, wie weit die deutlich einfachere Lösung der ungewichteten Summe trägt.

5 Ergebnisse

Nachdem auf den vorausgehenden Seiten ausführlich beschrieben wurde, wie Handlungsalternativen und -konsequenzen ermittelt werden können, und ein Vorschlag zur Messung und Skalierung der Nutzen bzw. subjektiven Wahrscheinlichkeiten unterbreitet wurde, soll nun kurz auf die Ergebnisse der Messung eingegangen werden. Hierfür wird zunächst kurz die Datenbasis der Untersuchung dargestellt. Es folgt eine Darstellung der allgemeinen Messergebnisse, eine Überprüfung der dimensional Struktur und schließlich eine Diskussion der Erklärungskraft der RC-Variablen.

5.1 Datenbasis

Die folgenden Analysen basieren auf einer postalischen Befragung von 1.500 konventionellen und 1500 ökologisch wirtschaftenden Landwirten aus Hessen, NRW und Niedersachsen. Die Stichprobe der konventionellen Landwirte wurde aus INVEKOS-Daten zur EU-Förderung gezogen. Die zweite Gruppe – Ökolandwirte – wurde als Stichprobe aus der Kartei der Öko-Kontrollbehörden der Bundesländer gezogen.

Grundlage waren alle landwirtschaftlichen Betriebe, die in den Jahren 2000, 2001 und 2002 zur Kontrolle nach Verordnung (EWG) Nr. 2092/91 über den ökologischen Landbau angemeldet wurden.⁷

Die Befragung wurde Anfang des Jahres 2004 durchgeführt; insgesamt gingen ausgefüllte Fragebögen von 969 Ökolandwirten und von 826 konventionellen Landwirten bei der Universität zu Köln ein, sodass eine (korrigierte) Ausschöpfungsquote von 63% erreicht wurde.⁸ Für die Zwecke dieser Untersuchung wurde die Stichprobe auf eine Teilpopulation eingegrenzt (siehe Best 2006 für eine Darstellung der Gesamtstichprobe): In der Gruppe der Ökolandwirte wurden nur Betriebe betrachtet, die zwischen 2000 und 2002 auf ökologische Landwirtschaft umgestellt wurden (d. h., Betriebe, die aufgrund von Änderungsmeldungen in die Stichprobe gerieten, wurden ausgeschlossen). Damit sollte zum einen eine systematische Verzerrung der Stichprobe vermieden, zum anderen eine valide Messung der entscheidungsbezogenen Variablen ermöglicht werden. Die Gruppe der konventionellen Landwirte wurde auf Betriebsleiter begrenzt, die angaben, schon einmal über eine Umstellung auf ökologische Landwirtschaft nachgedacht zu haben. Hierdurch konnte sichergestellt werden, dass alle Befragten sich tatsächlich in einer Entscheidungssituation befanden (die Ökolandwirte haben sich für, die konventionellen Landwirte gegen eine Umstellung entschieden), und valide Angaben zum Entscheidungsprozess machen können. Insgesamt reduzierte sich die Fallzahl durch die Begrenzungen auf 163 konventionell und 494 ökologisch wirtschaftende Landwirte.

- 7 Alle landwirtschaftlichen Betriebe, die ihre Produkte als ‚ökologisch‘ oder ‚biologisch‘ vermarkten wollen, müssen zum Kontrollverfahren nach der genannten Verordnung gemeldet sein. Diese ‚Meldung zum Kontrollverfahren‘ kann in zwei Kategorien unterschieden werden: ‚Neumeldungen‘ und ‚Änderungsmeldungen‘. Eine Neumeldung erfolgt, wenn der Betrieb auf ökologische Landwirtschaft umgestellt wird. Eine Änderungsmeldung wird hingegen bei Betrieben notwendig, die bereits ökologisch wirtschaften, aber meldepflichtige Änderungen vornehmen. Beispiele für solche Änderungen sind der Wechsel der Kontrollstelle oder wenn der Betrieb den ‚Kontrollbereich‘ nach Verordnung (EWG) Nr. 2092/91, Anhang III (vgl. EC 1991) wechselt (z. B. einen Hofladen eröffnet: Zum Bereich der ‚Erzeugung‘ kommt dann der Bereich ‚Vermarktung‘ hinzu.). Eine Unterscheidung zwischen den beiden Arten der Meldung konnte bei der Stichprobenziehung nicht getätigt werden, das Datum der Umstellung des Betriebes wurde jedoch in der Befragung erhoben.
- 8 Etwa 1,4% der 3.000 Fragebögen konnten wegen fehlerhafter Adressen nicht zugestellt werden und weitere 3,9% der Befragten waren aufgrund von Krankheit nicht befragbar oder nicht (mehr) Teil der Grundgesamtheit. Hierunter fallen insbesondere Landwirte, die ihren Hof aufgegeben haben bzw. in Rente sind, und solche, die nur noch für ihren Eigenbedarf wirtschaften, außerdem sozialpädagogische Einrichtungen und Mitglieder der Ökostichprobe, die ihren Hof bereits als Ökobetrieb übernommen haben (und als solche keine Angaben zum Umstellungsprozess machen konnten). Zieht man diese stichprobenneutralen Ausfälle von der Bruttostichprobe ab, bleibt eine bereinigte Bruttostichprobe von 2.843 Personen – hiervon 1.428 in der Ökostichprobe und 1.415 in der Vergleichsstichprobe.

5.2 Allgemeine Ergebnisse der Messung

Die Entscheidung zwischen zwei Handlungsalternativen beruht nach der SEU-Theorie (wie oben ausgeführt) auf einem Vergleich des Nettonutzens der Alternativen. Um diesen Vergleich empirisch nachvollziehen zu können, müssen für jede Handlungskonsequenz der Nutzen U und die subjektive Eintrittswahrscheinlichkeit p bestimmt werden. In den folgenden Absätzen werden die Ergebnisse der empirischen Messung dargestellt und diskutiert.

Tabelle 1 Nutzen und Wahrscheinlichkeiten der Handlungskonsequenzen

Konsequenz	Ökostichprobe			Vergleichsstichprobe		
	U^a	P_k^b	P_o^c	U^a	P_k^b	P_o^c
a Einfache und effektive Bekämpfung von Unkraut und Schädlingen	0,10	0,71	0,35	1,38	0,66	0,31
b Gute Preise für die Produkte	0,24	0,33	0,66	1,51	0,33	0,42
c Hoher Ertrag an landwirtschaftlichen Produkten	-0,07	0,61	0,33	1,02	0,64	0,23
d „Papierkram“ erledigen müssen	-0,55	0,65	0,73	-0,64	0,82	0,84
e Gesicherter Absatz der Produkte	0,26	0,55	0,63	1,19	0,52	0,40
f Abhängigkeit von Subventionen	-0,51	0,69	0,67	-1,30	0,70	0,74
g Sicherheit vor Lebensmittelskandalen	0,33	0,33	0,63	1,10	0,40	0,43
h Umweltfreundliche Produktionsweise	0,85	0,39	0,87	1,10	0,65	0,73
i Gutes Image als Landwirt in der Bevölkerung	0,54	0,42	0,72	1,29	0,52	0,64
j Ausreichend Freizeit	-0,37	0,34	0,34	0,68	0,34	0,21
k Hohe Prämien/Zuschüsse	0,22	0,36	0,68	-0,65	0,31	0,48
l Keine chemischen Spritzmittel verwenden	0,87	0,35	0,86	-0,12	0,36	0,80
m Umbauten an den Stallungen vornehmen	0,00	0,44	0,57	-0,11	0,51	0,66
n Langfristige Sicherung des Fortbestehens des Betriebes	0,28	0,40	0,57	1,17	0,50	0,39

N (Öko) = 494; N (Vergleich) = 163

^a Nutzen/Kosten der Handlungskonsequenz.

^b Wahrscheinlichkeit, dass die Konsequenz eintritt, wenn der Hof konventionell bewirtschaftet wird.

^c Wahrscheinlichkeit, dass die Konsequenz eintritt, wenn der Hof ökologisch bewirtschaftet wird.

Tabelle 1 stellt für alle Konsequenzen die Mittelwerte der Nutzen und Wahrscheinlichkeiten vergleichend dar. Wie man sehen kann, werden Kosten und Nutzen der einzelnen Konsequenzen im Großen und Ganzen so bewertet, wie man es ad hoc erwarten würde. Beispielsweise ist für konventionelle Landwirte einfache Unkraut-

bekämpfung, gute Preise und hoher Ertrag mit hohem Nutzen verbunden, während dies unter Ökolandwirten nicht so stark ausgeprägt ist. Nahezu alle Vorzeichen sind plausibel, was als Hinweis auf eine valide Messung gedeutet werden kann. Eine Ausnahme stellt lediglich der negative Nutzen dar, den Ökolandwirte ‚ausreichend Freizeit‘ zuschreiben. Der leicht negative Nutzen eines hohen Ertrages kann evtl. damit erklärt werden, dass ein hoher Ertrag mit Maßnahmen verbunden ist, die als umweltschädlich angesehen werden können, wie z. B. starke Düngung und Einsatz von Herbiziden und Pestiziden. Es zeigt sich außerdem, dass die Nutzen keineswegs über alle Akteure konstant sind, sondern sich systematisch zwischen den Gruppen unterscheiden.

Auch bezüglich der Wahrscheinlichkeiten kann festgestellt werden, dass diese zwischen Konsequenzen, Alternativen und Gruppen streuen. Die Streuung zwischen den Gruppen ist jedoch geringer als im Fall der Konsequenzen.⁹ Wie die Nutzen erscheinen auch die subjektiven Wahrscheinlichkeiten im Allgemeinen plausibel. So erwarten Angehörige beider Gruppen mit einer recht hohen Wahrscheinlichkeit von ca. 0,7, dass die Unkrautbekämpfung bei konventioneller Wirtschaftsweise einfach ist. Im Falle einer Umstellung auf Ökolandbau wird dies nur mit einer Wahrscheinlichkeit von etwa 0,3 erwartet. Um ein weiteres Beispiel zu nennen: Die Befragten schätzen die Wahrscheinlichkeit, von Subventionen abhängig zu sein, durchaus realistisch für beide Wirtschaftsweisen als recht hoch ein (um 0,7). Obwohl die Unterschiede hinsichtlich der subjektiven Wahrscheinlichkeiten zwischen den Gruppen vergleichsweise gering sind, können auch hier Differenzen beobachtet werden: Beispielsweise erwarten die Befragten der Ökostichprobe vom Ökolandbau eine deutlich umweltfreundlichere Produktionsweise, während in der Vergleichsstichprobe beide Alternativen ähnlich bewertet werden (d. h. konventionelle Landwirte schätzen den konventionellen Landbau deutlich umweltfreundlicher ein als Ökolandwirte). Weitere, allesamt plausible Unterschiede zwischen den Stichproben können z. B. für den Absatz oder die langfristige Sicherung des Betriebes konstatiert werden. Insgesamt kann damit auch bei den Wahrscheinlichkeiten von einer validen Messung ausgegangen werden.

Im Rahmen der RCT sind Nutzen und Wahrscheinlichkeiten jedoch für sich genommen nicht ausreichend für eine Entscheidung, sondern nur ein notwendiger Schritt zur Berechnung des Nettonutzens, der ja aus der Produktsomme dieser beiden Faktoren besteht. Tabelle 2 stellt daher die Produkte $p_i U_i$ (also den Beitrag der einzelnen Konsequenzen zum Nettonutzen der Alternative) und die Nutzendiffe-

9 Dies kann unter Umständen bedeuten, dass der Nutzen U für die Erklärung von Verhalten wichtiger ist als die subjektiven Wahrscheinlichkeiten p .

renzen der einzelnen Konsequenzen vergleichend zusammen. Da bereits bezüglich der in die Produkte eingehenden Nutzen und Wahrscheinlichkeiten festgestellt wurde, dass die Messung als valide angesehen werden kann, überrascht es nicht, dass sich auch die Nettonutzen und Nutzendifferenzen als sinnvoll interpretierbar erweisen.¹⁰ Bis auf wenige Ausnahmen weisen die Nutzendifferenzen in der Ökostichprobe ein positives Vorzeichen bzw. in der Vergleichsstichprobe ein negatives Vorzeichen auf oder zeigen im Größenvergleich zwischen den beiden Stichproben in die erwartete Richtung. Dies ist lediglich bei einzelnen Konsequenzen nicht (Preise) oder nur in geringem Maße (Subventionsabhängigkeit, Image und Umbauten) gegeben. Als Erklärung hierfür könnte dienen, dass Landwirte beispielsweise die Preise unabhängig von der Produktionsweise als zu niedrig und ihre Subventionsabhängigkeit als zu hoch ansehen. Eine systematische Variation zwischen den Handlungsalternativen wäre dann nicht zu erwarten.

Wie man sehen kann, weisen biserielle Korrelationen zwischen den Differenzen ND_i und der Gruppenzugehörigkeit bei nahezu allen Konsequenzen das erwartete positive Vorzeichen auf. Im Fall der Konsequenzen Preise, Papierkram, Subventionsabhängigkeit, Image und Umbauten (ND_b , ND_d , ND_r , ND_i und ND_m) ist die Korrelation so gering, dass davon ausgegangen werden kann, dass diese Konsequenzen keinen Einfluss auf die Entscheidung haben. Es handelt sich insofern um nicht-saliente Konsequenzen, die lediglich durch Ungenauigkeiten der qualitativen Methode in die Befragung gelangten. Andererseits wird sich dies nie vermeiden lassen, wenn man das Risiko minimieren will, relevante Handlungskonsequenzen zu vergessen. *Die nicht-salienten Konsequenzen werden für die Konstruktion der Gesamt-Nettonutzen nicht verwendet und im Folgenden daher nicht weiter betrachtet.*

An dieser Stelle soll noch auf ein Problem hingewiesen werden, das sich aus der Messung auf 5-stufigen Ratingskalen und der Bildung von Produkttermen ergibt. Die Teil-Erwartungsnutzen der einzelnen Konsequenzen wie auch die einzelnen Differenzen weisen eine große Zahl von Nullen auf (zwischen 36% bei ND_h und 71% bei ND_j). Dies ist zwar ausgesprochen unangenehm, kann jedoch bei dieser Art der Messung nicht vermieden werden: Bei zufälliger Beantwortung der Items (d. h. der Befragte tippt blind) erhält man mit einer Wahrscheinlichkeit von 20% einen Nutzen von Null und mit einer ebenso hohen Wahrscheinlichkeit gleiche Wahrscheinlichkeiten für beide Alternativen, also eine Differenz von Null. Da eine Null

10 Bei Befragten mit einzelnen fehlenden Angaben wurde der Nettonutzen durch die Zahl der gültigen Angaben dividiert und mit neun multipliziert. Dadurch hat der Nettonutzen bei allen Befragten die gleiche Spannweite.

bei Multiplikationen dominant ist, erhält man bei zufälligem Ausfüllen der Itembatterien in 40% der Fälle eine ND_i von Null. Die hohe Anzahl der Nullen ist damit rein artifiziell (und kann a posteriori nicht mehr durch Recodierungen oder ähnliches verändert werden). Es wäre sinnvoll, in zukünftigen empirischen Anwendungen der RCT zumindest 7-stufige Ratingskalen zu verwenden – in diesem Fall wären nur 28,5% Nullen im Produkt zu erwarten. Eine noch günstigere Alternative, die jedoch sehr aufwendig und nur im persönlichen Interview zu verwirklichen ist, wäre die Verwendung einer Magnitude-Skalierung (siehe Lodge 1981).

Tabelle 2 Teil-Erwartungsnutzen und ND_i der Handlungskonsequenzen

Konsequenz	Ökostichprobe			Vergleichsstichprobe			
	NN_k^a	NN_o^b	ND^c	NN_k^a	NN_o^b	ND^c	r_{bis}^d
a Einfache und effektive Bekämpfung von Unkraut und Schädlingen	0,12	0,08	-0,03	0,94	0,38	-0,56	0,47
b Gute Preise für die Produkte	0,09	0,18	0,08	0,51	0,64	0,13	-0,03
c Hoher Ertrag an landwirtschaftlichen Produkten	-0,02	0,01	0,03	0,68	0,22	-0,45	0,51
d „Papierkram“ erledigen müssen	-0,36	-0,39	-0,04	-0,49	-0,55	-0,05	0,04
e Gesicherter Absatz der Produkte	0,16	0,21	0,05	0,62	0,47	-0,15	0,35
f Abhängigkeit von Subventionen	-0,37	-0,35	0,02	-0,92	-0,97	-0,05	0,08
g Sicherheit vor Lebensmittel-skandalen	0,12	0,27	0,15	0,46	0,50	0,05	0,12
h Umweltfreundliche Produktionsweise	0,35	0,78	0,43	0,74	0,82	0,08	0,27
i Gutes Image als Landwirt in der Bevölkerung	0,25	0,46	0,20	0,69	0,84	0,16	0,05
j Ausreichend Freizeit	-0,04	-0,02	0,02	0,31	0,16	-0,15	0,29
k Hohe Prämien/Zuschüsse	0,12	0,23	0,11	-0,20	-0,32	-0,12	0,26
l Keine chemischen Spritzmittel verwenden	0,31	0,79	0,48	0,06	-0,05	-0,11	0,41
m Umbauten an den Stallungen vornehmen	0,04	0,02	-0,02	0,02	-0,05	-0,06	0,05
n Langfristige Sicherung des Fortbestehens des Betriebes	0,15	0,24	0,09	0,64	0,49	-0,15	0,33

N (Öko) = 494; N (Vergleich) = 163

^a Erwartungsnutzen der Konsequenz, wenn der Hof konventionell weiterbewirtschaftet wird.

^b Erwartungsnutzen der Konsequenz, wenn der Hof auf Ökolandbau umgestellt wird.

^c Differenz der Teil-Erwartungsnutzen der Konsequenz ($NN_o - NN_k$).

^d Biserielle Korrelation zwischen ND_i und Gruppenzugehörigkeit (ungewichtet).

5.3 Dimensionale Struktur und Summenbildung

Friedrichs et al. (1993) regen an, die Konsequenzen auf ihre dimensionale Struktur hin zu prüfen. Falls sich eine klare dimensionale Struktur empirisch nachweisen lässt (die RCT macht hierzu keine Aussage), bestünde die Gefahr, dass man bei einer reinen Summenbildung einzelne Dimensionen über- oder unterrepräsentiert.

Die dimensionale Struktur wurde durch eine Hauptkomponentenanalyse empirisch geprüft. Tabelle 3 stellt die Ergebnisse der Faktorextraktion nach dem Kaiserkriterium dar (Faktorladungen und Kommunalitäten). Wie man der Tabelle entnehmen kann, ist keine klare dimensionale Struktur der Konsequenzen zu erkennen. In der Ökostichprobe werden zwei Faktoren extrahiert, die sich – wenn auch mit etwas Mühe – benennen ließen: Der erste Faktor könnte mehr oder weniger für umweltbezogene Folgen, der zweite für ökonomische und betriebliche Folgen stehen. In der Vergleichsstichprobe ergibt sich eine dreidimensionale Lösung: Auch hier könnte der erste Faktor (F1) für ökonomische Konsequenzen stehen und ein weiterer (F3) teilweise für Umwelt, der verbleibende zweite Faktor wäre am ehesten einer der betrieblichen Folgen. Insgesamt ist die Faktorenstruktur jedoch uneinheitlich und nicht befriedigend. Selbst Faktoren, die hier gleich benannt wurden, entsprechen einander nicht. Würde man sie trotz dieser Schwierigkeiten als bedeutungsvolle Dimensionen interpretieren, wäre dennoch nicht zu befürchten, dass eine einzelne Dimension überrepräsentiert ist: Auf allen Faktoren laden vier (in einem Fall fünf) Konsequenzen höher als 0,3, sodass alle Dimensionen in der Summenskala etwa gleich stark repräsentiert wären. Insgesamt erscheint es aber gerechtfertigt, die Konsequenzen als unabhängig voneinander zu interpretieren.

Der Nettonutzen NN der beiden Alternativen ‚Ökolandbau‘ und ‚weiter wie bisher‘ wird daher als (ungewichtete) Summe von Nutzen und Wahrscheinlichkeiten der neun Konsequenzen Unkrautbekämpfung, Ertrag, Absatz, Sicherheit vor Skandalen, umweltfreundliche Produktion, Freizeit, Prämien, keine Chemie und langfristige Absicherung gebildet. Für Befragte, die Angaben zu weniger als sechs der neun Konsequenzen gemacht haben, werden keine Nettonutzen berechnet.

Die zentrale RC-Variable ist die Nettonutzendifferenz ND, die als mathematische Beschreibung des Vergleichs zweier Handlungsalternativen angesehen wird. Sie wird als $ND = NN_{\text{öko}} - NN_{\text{konv}}$ berechnet und kann – bei der vorliegenden Skalierung – theoretisch Werte zwischen -18 und +18 annehmen.¹¹ Empirisch liegt

11 Diese Extremwerte würden nur erreicht, wenn ein Befragter alle Nutzenitems mit ‚sehr gut‘ oder alle mit ‚sehr schlecht‘ bewerten würde und die Wahrscheinlichkeiten für die Alternative ‚Ökolandbau‘ alle mit 1, die Wahrscheinlichkeiten für ‚weiter so‘ dagegen alle mit 0 bewerten würde (oder umgekehrt).

die Nettotonutzendifferenz zwischen -8,5 und +9,5. Eine positive Nutzendifferenz verweist auf Vorteile des Ökolandbaus.

Tabelle 3 Hauptkomponentenanalyse der neun Konsequenzen (Varimax-Rotiert)

Konsequenz	Ökostichprobe			Vergleichsstichprobe			
	F 1	F 2	h ²	F 1	F 2	F 3	h ²
ND _a Einfache und effektive Bekämpfung von Unkraut und Schädlingen		0,76	0,59	0,72			0,52
ND _c Hoher Ertrag an landwirtschaftlichen Produkten		0,73	0,54	0,53	0,41		0,45
ND _e Gesicherter Absatz der Produkte		0,38	0,19		0,76		0,58
ND _g Sicherheit vor Lebensmittelskandalen	0,67		0,49			0,78	0,66
ND _h Umweltfreundliche Produktionsweise	0,82		0,70			0,71	0,56
ND _j Ausreichend Freizeit	0,32		0,13		0,80		0,66
ND _k Hohe Prämien/Zuschüsse		0,57	0,33	0,58		-0,32	0,44
ND _l Keine chemischen Spritzmittel verwenden	0,70		0,49	0,66			0,44
ND _n Langfristige Sicherung des Fortbestehens des Betriebes	0,64		0,42	0,48	0,38	0,45	0,58
Eigenwert	2,72	1,15		2,21	1,45	1,24	

$N(\text{Öko}) = 427; N(\text{Vergleich}) = 148$

5.4 Erklärungskraft der RC-Variablen

Wie bereits in Abschnitt 4.1 dargelegt, kann das Ergebnis der Messung als plausibel angesehen werden, sodass zumindest prinzipiell von einer validen Messung ausgegangen werden kann. In diesem Abschnitt soll als zusätzliche Validitätsprüfung, und um den Nutzwert einer direkten Anwendung der RCT aufzuzeigen, noch die Prognosegüte der RC-Variablen untersucht werden. Wenn die RCT zutreffend ist, sollte eine Reihe von Tatbeständen auch empirisch zu finden sein: Eine Mindestforderung ist, dass Ökolandwirte, die sich ja für eine Umstellung entschieden haben, der ökologischen Landwirtschaft im *Durchschnitt* einen höheren Nettotonutzen zuweisen als der konventionellen Landwirtschaft. Die Nettotonutzendifferenz sollte zudem in multivariaten Modellen einen statistisch signifikanten und zumindest mäßig starken Einfluss auf die Zielvariable haben. Nimmt man eine deterministische Variante der RCT als gültig an, sollte sich mit der Nettotonutzendifferenz nicht nur

die Wahrscheinlichkeit einer Umstellung erhöhen, sondern *alle* Landwirte, die der ökologischen Landwirtschaft einen höheren Nutzen zuweisen, sollten ihren Betrieb umstellen, aber dagegen *keiner* der Landwirte mit einer negativen Nettonutzendifferenz.

Tabelle 4 Klassifikation anhand der Nettonutzendifferenz

Nettonutzendifferenz	Konventionelle Landwirte %	Ökolandwirte %	N
negativ	76,1	28,4	255
$-0,5 \leq ND \leq 0,5$	13,2	11,4	75
positiv	10,7	60,2	301
N (100%)	159	472	

Tabelle 4 kann man entnehmen, dass die deterministische, sehr restriktive Variante der RCT mit den vorliegenden Daten, wenig überraschend, nicht bestätigt werden kann: Es gibt auch konventionelle Landwirte, die der ökologischen Landwirtschaft einen höheren Nutzen zuweisen und Ökobauern, die nach der reinen Lehre ihren Betrieb konventionell betreiben müßten. Überraschend ist jedoch das geringe Ausmaß dieser Fehlklassifikationen: Nur etwa 10% der konventionellen und unter 30% der ökologisch wirtschaftenden Landwirte haben eine unpassende Nettonutzendifferenz. An anderer Stelle (siehe Best 2006) wird zudem gezeigt, dass die Fehlklassifikation der Ökolandwirte auf den Einfluss von Umwelteinstellungen zurückgeführt werden kann.

Die methodologisch etwas weniger anspruchsvolle, in der Sozialwissenschaft jedoch weitaus üblichere – und wohl auch angemessenere – Testmethode meistern die RC-Variablen ohne Probleme: Konventionelle Landwirte erwarten im Mittel von der konventionellen Landwirtschaft einen höheren Nutzen als von einer Umstellung auf ökologische Landwirtschaft, Ökobauern erwarten dagegen von der ökologischen Landwirtschaft durchschnittlich einen höheren Nutzen als von der konventionellen Landwirtschaft. Entsprechend haben konventionelle Landwirte eine negative, Ökobauern eine positive Nettonutzendifferenz (siehe Tabelle 5). Die Ergebnisse sind statistisch hochsignifikant. Interessant, aber für die Zwecke dieses Beitrages nicht von Belang, ist die Tatsache, dass die Nettonutzendifferenz im wesentlichen einer unterschiedlichen Bewertung der konventionellen Landwirtschaft, nicht aber einer unterschiedlichen Bewertung der Ökolandwirtschaft geschuldet ist. Dieses Resultat verweist darauf, dass im Untersuchungszeitraum von 2000 bis 2002 viele Landwirte, die sich letztendlich für eine Umstellung auf ökologische Land-

wirtschaft entschieden haben, mit der konventionellen Landwirtschaft hochgradig unzufrieden waren. Vor dem Hintergrund der damals aktuellen BSE-Krise und dem schlechten Image der (konventionellen) Landwirtschaft in der Bevölkerung sollte dies nicht überraschen.

Tabelle 5 Nettonutzen und Nettonutzendifferenz nach Wirtschaftsweise

	Konventionelle Landwirte		Ökolandwirte		t
	\bar{x}	σ	\bar{x}	σ	
NN_o	2,70	2,64	2,61	3,37	0,3
NN_k	4,25	2,43	1,26	2,39	-13,6***
ND	-1,57	2,06	1,35	2,91	13,8***
N_{min}	160		472		

+: $p \leq 0,1$; *: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$ (zweiseitig).

^a NN_o = Nettonutzen Öko, NN_k = Nettonutzen konv, ND = Nettonutzendifferenz.

Wie aufgrund der Ergebnisse zu erwarten, besteht eine starke Korrelation zwischen der Nettonutzendifferenz und der Wahrscheinlichkeit einer Umstellung. Bivariat ergibt sich ein Zusammenhang von $r_{bis} = 0,58$. Dieser bivariate Zusammenhang bleibt auch unter Kontrolle sozio-demografischer und hofstruktureller Merkmale (z. B. Rinderhaltung, Nebenerwerb, Nutzfläche) stabil: In einer multiplen logistischen Regression liegt der xy^* -standardisierte Logit-Koeffizient der Nettonutzendifferenz bei 0,54.¹² Die sich aus einer direkten Anwendung der RCT ergebenden Variablen erlauben damit eine (Pseudo-)Varianz-Erklärung, die weit über dem Erwartungswert sozialwissenschaftlicher Untersuchungen liegt. Dieses Ergebnis verweist zum einen auf die Stärke der RCT, zum anderen auf Vorteile einer direkten Anwendung.

Bei der Interpretation der Ergebnisse sollte jedoch beachtet werden, dass die Berechnungen anhand von Daten einer Querschnittsbefragung erfolgten. Daher wurden die Landwirte retrospektiv befragt, wenn auch das retrospektive Zeitfenster stark begrenzt wurde, um valide Antworten sicher zu stellen. Dennoch ist es möglich, dass hierbei Verzerrungen, beispielsweise durch nachträgliche Rationalisierungen auftreten, und eine Kausalinterpretation erschweren. In zukünftigen Anwendungen der RCT wäre es daher günstig, wenn auch sicher nicht billig, auf Panel-Befragungen zurückzugreifen.

12 Vgl. Long (1997: 69ff.) zur Standardisierung von Koeffizienten in der logistischen Regression. Die Standardisierung bezieht sich auf die gemessene x-Variable und eine geschätzte, latente Variable y^* , die – so die Annahme – der beobachteten dichotomen Variable y zugrunde liegt.

6 Zusammenfassung

Ausgehend von der Feststellung, dass die direkte empirische Anwendung der Theorie rationalen Handelns bislang in der Sozialwissenschaft nur unzureichend Aufmerksamkeit erfahren hat, wurde in diesem Beitrag ein Vorschlag zur Operationalisierung von RC-Variablen vorgestellt.

Es wurde vorgeschlagen, die Handlungsalternativen, für die Daten erhoben werden, fest vorzugeben. Die Zahl der Alternativen sollte zudem stark beschränkt werden – idealerweise auf zwei. Hierdurch soll die Belastung der Befragten möglichst gering gehalten werden. Um Verzerrungen durch eine verdeckte dritte Alternative auszuschließen, können weitere Alternativen in einer verkürzten Form abgefragt werden. Die Konsequenzen, die sich aus einer Entscheidung ergeben, müssen empirisch ermittelt werden. Hier erweist es sich als empfehlenswert, in vorgeschalteten qualitativen Interviews die Konsequenzen zu ermitteln, um diese dann in der Hauptbefragung geschlossen vorzugeben. Für jede der so ermittelten Konsequenzen muss auf einer Rating-Skala der Nutzen U_i und getrennt für beide Handlungsalternativen jeweils die subjektive Eintrittswahrscheinlichkeit p_i ermittelt werden. Um mathematische Probleme zu vermeiden ist es empfehlenswert, die Zahl der Kategorien hierbei eher hoch zu wählen (sieben oder mehr, selbst wenn dies Zufallsvarianz erzeugt). Die so gemessenen Nutzen sollten bipolar codiert werden, die Wahrscheinlichkeiten unipolar zwischen Null und eins. Es lässt sich argumentieren, dass diese Messung und Codierung unter bestimmten Voraussetzungen (insbesondere einer Gleichabständigkeit der Antwortkategorien) zu Variablen auf Ratioskalenniveau führt. Dieses Skalenniveau ist wiederum Voraussetzung für eine unkomplizierte Verwendung der Produktterme in Regressionen.

Am Beispiel einer Studie zur Umstellung auf ökologische Landwirtschaft wurde gezeigt, dass dieses recht einfache Verfahren zu eindeutigen Ergebnissen führen kann. Die Bewertung der Konsequenzen (also die Teil-Erwartungsnutzen) lassen Rückschlüsse darauf zu, welche Teilaspekte für die Entscheidung besonders relevant sind, sodass eine Vielzahl von Informationen zum Studienobjekt – hier also der Umstellung auf ökologische Landwirtschaft – gewonnen werden kann. Der Nettonutzen der Alternativen (bzw. die Nutzendifferenz) ist sehr hoch mit der Wahrscheinlichkeit einer Umstellung korreliert, sodass von einer validen Messung ausgegangen werden kann. Es wurde gezeigt, dass anhand der RC-Variablen eine sehr gute Varianzaufklärung geleistet werden kann.

Es bleibt zu wünschen, dass dieser Beitrag einen praktikablen Vorschlag zur Operationalisierung der RCT präsentiert. Die direkte Anwendung der RCT bietet eine ganze Reihe von Vorteilen. Es wurde bereits darauf verwiesen, dass bei dieser Vari-

ante der empirischen RCT-Anwendung eine Vielzahl von Informationen zum Studienobjekt gewonnen werden, unter anderem zur relativen Wichtigkeit verschiedener Handlungskonsequenzen. Diese Informationen werden jedoch vor dem Hintergrund einer allgemeinen Handlungstheorie gewonnen und interpretiert, sodass die Beliebbarkeit nicht-theoriegeleiteter Fragebogen-Items vermieden werden kann. Im Vergleich zur rein heuristischen Anwendung von RC können zudem Annahmen über die Wirkung verschiedener Randbedingungen detailliert überprüft werden, anstatt lediglich ihre Gültigkeit vorauszusetzen. Auch die Frage, ob das SEU-Modell mit seiner Entscheidungsregel realen menschlichen Entscheidungen gerecht wird, lässt sich in direkten empirischen Anwendungen besser und genauer überprüfen als im Modellbildungs-Paradigma. Von einer stärkeren Verankerung der RCT in der empirischen Sozialforschung würde sicherlich sowohl die Theorieentwicklung als auch die empirische Forschung profitieren.

Literatur

- Ajzen, I., 1991: The Theory of Planned Behavior, *Organizational Behavior and Human Decision Processes* 50: 179-211.
- Ajzen, I. und M. Fishbein, 1980: *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs: Prentice-Hall.
- Allison, P.D., 1977: Testing for Interaction in Multiple-Regression, *American Journal of Sociology* 83: 144-153.
- Best, H., 2006: *Die Umstellung auf ökologische Landwirtschaft als Entscheidungsprozess*. Wiesbaden: VS-Verlag.
- Bioland, 2002: *Bioland-Richtlinien*. Mainz: Bioland e.V. Verband für organisch-biologischen Landbau.
- Bouffard, J.A., 2002: Methodological and Theoretical Implications of Using Subject-generated Consequences in Tests of Rational Choice Theory. *Justice Quarterly* 19: 747-771.
- Braun, N., und A. Franzen, 1995: Rationalität und Umweltverhalten. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47: 231-248.
- Brüderl, J., 2004: Die Überprüfung von Rational-Choice-Modellen mit Umfragedaten. S. 163-180 in: A. Diekmann und T. Voss (Hg.): *Rational-Choice-Theorie in den Sozialwissenschaften*. München: Oldenbourg.
- Coleman, J. S., 1995: *Grundlagen der Sozialtheorie, Band 1: Handlungen und Handlungssysteme*. München: Oldenbourg.
- Demeter, 2002: *Erzeugungsrichtlinien für die Anerkennung der Demeter-Qualität*. Darmstadt: Demeter-Bund e.V.
- Diekmann, A., 1996: Homo Öconomicus. S. 89-118 in: A. Diekmann und C. C. Jaeger (Hg.): *Umweltsoziologie*. Opladen: Westdeutscher Verlag.
- Diekmann, A., 2002: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek: Rowohlt.
- EC, 1991: Council Regulation (EC) No 2092/91/EEC of 24 June 1991 on organic production of agricultural products and indications referring thereto on agricultural products and foodstuffs. *Official Journal of the European Communities* 1991: 1-15.
- Eisenführ, F., und M. Weber, 1994: *Rationales Entscheiden*. Berlin: Springer.

- Evans, M.G., 1991: The Problem of Analyzing Multiplicative Composites. *Interactions Revisited*. *American Psychologist* 46: 6-15.
- Fishbein, M., und I. Ajzen, 1975: *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, Mass: Addison-Wesley.
- Fishburn, P., 1981: Subjective expected utility: A review of normative theories. *Theory and Decision* 13: 139-199.
- Friedman, M., 1953: The Methodology of Positive Economics. S. 3-43 in: M. Friedman (Hg.): *Essays in Positive Economics*. Chicago: Chicago University Press.
- Friedrichs, J., 1980: *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag.
- Friedrichs, J., M. Stolle und G. Engelbrecht, 1993: Rational-Choice-Theorie: Probleme der Operationalisierung. *Zeitschrift für Soziologie* 22: 2-15.
- Goldthorpe, John H., 1996: The Quantitative Analysis of Large-Scale Data-Sets and Rational Action Theory. For a Sociological Alliance. *European Sociological Review* 12: 109-126.
- IFOAM, 2002: *IFOAM Basic Standards for Organic Production and Processing*. Bonn: International Federation of Organic Agriculture Movements.
- Kahnemann, D. und A. Tversky, 1984: Choices, Values and Frames. *American Psychologist* 39: 341-350.
- Kelle, U. und C. Lüdemann, 1995: Grau, Teurer Freund, ist alle Theorie. *Rational Choice und das Problem der Brückenannahmen*. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47: 249-267.
- Kunz, V., 1994: Die empirische Prüfung von Nutzentheorien. S. 112-131 in: V. Kunz und U. Druwe (Hg.): *Rational Choice in der Politikwissenschaft*. Opladen: Leske+Budrich.
- Lampkin, N., 1994: *Organic Farming: Sustainable Agriculture in Practice*. S. 3-10 in: N. Lampkin und S. Padel (Hg.): *The Economics of Organic Farming. An International Perspective*. Wallingford: CAB-International.
- Lindenberg, S., 1996a: Die Relevanz Theoriereicher Brückenannahmen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48: 126-140.
- Lindenberg, S., 1996b: Theoriegesteuerte Konkretisierung der Nutzentheorie. Eine Replik auf Kelle/Lüdemann und Opp/Friedrichs. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48: 560-565.
- Lodge, M., 1981: *Magnitude Scaling. Quantitative Measurement of Opinions*. Beverly Hills: SAGE.
- Long, J. S., 1997: *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: SAGE.
- Lüdemann, C., 1992: Das Modell rationalen Handelns und der "deal" im Strafprozess. Ergebnisse einer empirischen Studie. *Zeitschrift für Rechtssoziologie* 13: 88-109.
- Mäder, P., A. Fließbach, D. Dubois, L. Gunst, P. Fried und U. Niggli, 2002: Soil fertility and biodiversity in organic farming. *Science* 296: 1694-1697.
- Mayntz, R., K. Holm, und P. Hübner, 1969: *Einführung in die Methoden der empirischen Soziologie*. Köln und Opladen: Westdeutscher Verlag.
- Naturland, 2002: *Naturland Richtlinien*. Gräfelting: Naturland - Verband für naturgemäßen Landbau e.V.
- Opp, K.-D., 1979: Das "ökonomische Programm" in der Soziologie. S. 313-350 in: H. Albert und J. H. Stapf (Hg.): *Theorie und Erfahrung. Beiträge zur Grundlagenproblematik der Sozialwissenschaften*. Stuttgart: Klett-Cotta.
- Opp, Karl-Dieter, 1990: Testing Rational Choice Theory in Natural Settings. S. 87-102 in: J.J. Hox und J. de Jong-Gierveld (Hg.): *Operationalization and Research Strategy*. Amsterdam: Swets & Zeitlinger.
- Opp, K.-D., 1999: Contending conceptions of the theory of rational action. *Journal of Theoretical Politics* 11: 171-202.
- Opp, K.-D., K. Burow-Auffahrt, P. Hartmann, T. von Witzleben, V. Pöhls und T. Spitzley, 1984: *Soziale Probleme und Protestverhalten*. Opladen: Westdeutscher Verlag.

- Opp, K.-D. und W. Roehl, 1990: Der Tschernobyl-Effekt. Eine Untersuchung über die Ursachen politischen Protests. Wiesbaden: Westdeutscher Verlag.
- Opp, K.-D. und J. Friedrichs, 1996: Brückenannahmen, Produktionsfunktionen und die Messung von Präferenzen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48: 546-559.
- Pratkanis, A. R., 1989: The cognitive representation of attitudes. S. 71-98 in: A.R. Pratkanis, S.J. Breckler und A.G. Greenwald (Hg.): *Attitude structure and function*. Hillsdale, NJ: Erlbaum.
- Rohrmann, B., 1978: Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie* 9: 222-245.
- Sarle, W.S., 1995: Measurement theory: Frequently asked questions, *Disseminations of the International Statistical Applications Institute* 1: 61-66.
- Scheuch, E.K. und H. Zehnpfennig, 1974: Skalierungsverfahren in der Sozialforschung. S. 97-203 in: R. König (Hg.): *Handbuch der empirischen Sozialforschung, Band 3a: Grundlegende Methoden und Techniken, Zweiter Teil*. Stuttgart: Ferdinand Enke Verlag.
- Schnell, R., P. Hill und E. Esser, 2005: *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- Smith, A., 1926: *Theorie der ethischen Gefühle*, 2. Band. Leipzig: Verlag von Felix Meiner.
- Tversky, A. und D. Kahnemann, 1988: Rational Choice and the Framing of Decisions. *Journal of Business* 59: 251-278.

Korrespondenzadresse: Henning Best
Lehrstuhl für Methoden der
empirischen Sozialforschung
und angewandte Soziologie
Fakultät für Sozialwissenschaften
Universität Mannheim
68131 Mannheim
best@uni-mannheim.de

Rezensionen



YVONNE KÜSTERS,
2006: Narrative
Interviews. Grundlagen
und Anwendungen.
Wiesbaden: VS Verlag.
ISBN: 978-3-531-15205-9,
204 Seiten, 16,90 EUR



ARND-MICHAEL NOHL,
2006: Interview und
dokumentarische Me-
thode. Anleitungen für
die Forschungspraxis.
Wiesbaden: VS-Verlag.
ISBN: 978-3-531-14983-7,
127 Seiten, 12,90 EUR.



UDO KUCKARTZ,
2007: Einführung in
die computergestützte
Analyse qualitativer
Daten. 2., aktuali-
sierte und erweiterte
Auflage. Wiesbaden:
VS-Verlag. ISBN: 978-
3-531-34247-4, 268
Seiten, 19,90 EUR.

Offene Interviews führen und auswerten *Sammelbesprechung*

Während seit Jahrzehnten sehr gute Lehrbücher in die Durchführung und Auswertung quantitativer Studien einleiten, kreiste der qualitative Methodendiskurs lange bis auf wenige Ausnahmen um (wissenschafts)theoretische Fragen, während die Forschungspraxis als Kunstlehre galt. Es gab keine andere Möglichkeit, die Techniken der qualitativen Datenerhebung und -auswertung zu erlernen, als bei einem erfahrenen Forscher an einem Beispiel. In den vergangenen Jahren versuchten zahlreiche Lehrbücher, diese Lücke zu schließen, unter anderem Yvonne Küsters' „Narrative Interviews“ (2006), Arnd-Michael Nohls „Interview und dokumentarische Methode“ (2006) und die mittlerweile in der dritten Überarbeitung erschienene „Einführung in die computergestützte Analyse qualitativer Daten“ von Udo Kuckartz (2007). Alle drei stammen aus sehr unterschiedlichen Forschungstraditionen und fokussieren auf unterschiedliche Auswertungsschritte.

Yvonne Küsters führt in ihrem Buch „Narrative Interviews“ (2006) in das von Fritz Schütze (1983) entwickelte Datenerhebungsverfahren der narrativen Interviews und das dazugehörige Auswertungsverfahren ein. „Bei dieser Befragungsform bittet man den Befragten mit einer Eingangsfrage um die Erzählung eines prozesshaften Vorgangs, den er selbst erlebt hat (wie z. B. seine Lebensgeschichte), ohne dass man seine Erzählung durch Nachfragen unterbricht“ (S. 13). Das Verfahren versucht so, das Problem der Reaktivität zu umgehen. Es basiert auf der Annahme, dass Stehgreiferzählungen beim Erzähler Strukturierungszwänge (Detailierungszwang, Gestaltschließungszwang sowie Relevanzfestlegungs- und Kondensierungszwang) erzeugen, in denen sich der Erzähler offenbart: „Auch vom Erzähler bewusst Verschwiegene und sogar Verdrängte wird

in der Erzählung erkennbar, denn es schlägt sich im Erzähltext als Zögern, in Lücken, in thematischen Brüchen, Wechseln der Textorte, Schweigen etc. nieder“ (S. 28).

Küsters erläutert den gesamten Forschungsprozess Schritt für Schritt jeweils erst allgemein, dann an einem konkreten Forschungsbeispiel aus der Musiksoziologie: Welche biographischen Entwicklungen führen dazu, dass ein Erwachsener, der noch nie ein Musikinstrument gespielt hat, damit beginnt, ein solches zu erlernen? In Kapitel 2 erläutert Küsters anhand dieses Beispiels, wie man eine Fragestellung entwickeln kann, was narrative Interviews sind und warum sie für diese Fragestellung das geeignete Datenerhebungsverfahren sind; wie sie ihre acht Interviewpartner ausgesucht hat und welche Probleme dabei aufgetreten sind.

Hierauf aufbauend, erläutert Küsters in Kapitel 3 die theoretischen und methodologischen Grundlagen des Verfahrens, das Schütze unter Bezugnahme auf Alfred Schütz, die Chicago School, den Symbolischen Interaktionismus, die Ethnomethodologie, die Konversationsanalyse und Grounded Theory entwickelte. Gemeinsam ist diesen Theorien das Ziel, dass sie soziale Prozesse untersuchen wollen, sowie die Annahme, „dass die soziale Wirklichkeit nicht außerhalb des Handelns der Gesellschaftsmitglieder existiert, sondern jeweils im Rahmen kommunikativer Interaktionen hergestellt wird. (...) Um die soziale Wirklichkeit zu untersuchen, müssen die kommunikativen Interaktionen sinnverstehend analysiert werden“ (S. 18). Die Herausforderung für die sozialwissenschaftliche Forschung besteht darin, dass Menschen zwar als Gesellschaftsmitglieder über ein Repertoire an Sinnbezügen, Deutungsmustern, Handlungsorientierungen und subjektiven Sichtweisen verfügen (müssen), um in sozialen Situationen überhaupt handeln und das Handeln anderer richtig deuten zu können, dass dies für sie jedoch so selbstverständlich ist, dass sie es nicht (vollständig) explizieren können. Die teilnehmende Beobachtung und offene Interviews sollen dem Forscher hel-

fen, indirekt auch auf den Teil sozialer Wirklichkeit zuzugreifen, der dem Handelnden nicht bewusst ist. In Auseinandersetzung mit Kritikern narrativer Interviews und anderen qualitativen und quantitativen Verfahren erläutert Küsters, für welche Fragestellungen sich narrative Interviews eignen und für welche nicht (Kapitel 3.2, 3.3 und 4.1).

In Kapitel 4.2 gibt Küsters Hinweise zur Auswahl der Interviewpartner und zum Einstieg ins Feld. Da die Methode sehr hohe Anforderungen an den Interviewer stellt, erklärt Küsters in Kapitel 4.3 für die verschiedenen Phasen des Interviews (Vorgespräch, Erzählstimulus, Aushandlungsphase, Haupterzählung, Koda, Nachfragephase, Nachgespräch, Interviewerprotokoll), welche typischen Fehler unterlaufen können und wie diese zu umgehen sind. Die Phase des Nachfragens erlaubt dem Interviewer, für seine Forschungsfrage relevante, aber in der Erzählung noch nicht genannte Informationen nachträglich einzuholen. Immanente Nachfragen zielen „auf bisher Unerzähltes, das aber in der Haupterzählung doch repräsentiert war“ (S. 61). Über exmanente Nachfragen kann der Interviewer „selber Themen ins Interview einbringen und den Erzähler auch zu Beschreibungen und Argumentationen auffordern“ (S. 63). Da später bei der Auswertung auch darauf geachtet werden soll, was der Erzähler verschweigen wollte, ist eine sorgfältige und aufwändige Transkription erforderlich (Kapitel 4.4.1).

In Kapitel 4.4.4 nennt Küsters Verfahren, mit denen narrative Interviews ausgewertet werden können. Die beiden wichtigsten sind die von Schütze entwickelte Narrationsstrukturanalyse (Kapitel 4.4.2, 4.5 und 4.6) und die biographische Fallrekonstruktion von Rosenthal (Kapitel 4.4.3). Küsters konzentriert sich auf eine ausführliche Erläuterung der von Schütze für die Auswertung vorgeschlagenen Arbeitsschritte: „1. Formale Textanalyse, 2. strukturelle inhaltliche Beschreibung der Darstellungsstücke, 3. analytische Abstraktion, 4. Wissensanalyse, 5. kontrastive Vergleiche unterschiedlicher Interviewtexte,

6. Konstruktion eines theoretischen Modells (...). Die Schritte 1. bis 4. dienen der Analyse eines Falls, also eines Interviews, während die Schritte 5. und 6. die Einzelanalysen aller Interviews eines Samples zusammenführen.“ (S. 77).

Das Buch schließt mit einem Überblick, in welchen Forschungsfeldern bisher narrativ gearbeitet wurde (Kapitel 5), einer Diskussion der Frage, ob und wie narrative Interviews auch außerhalb Deutschlands eingesetzt werden können (Kapitel 6), sowie mit praktischen Tipps zum Einstieg in die Forschungspraxis.

Das Buch besticht durch seine Anschaulichkeit und große Praxishöhe. Besonders hilfreich ist, dass an einem konkreten Forschungsbeispiel auf typische Fallstricke und Fehlerquellen hingewiesen wird. Während sich etwa Helfferichs (2005) Einführung in die qualitative Interviewführung an Dozenten richtet, eignet sich Küsters deshalb zum Selbststudium, sollte aber idealerweise durch ein Buch zum theoretischen Hintergrund der Biographieforschung ergänzt werden, etwa Fuchs-Heinritz (2005). Unbefriedigend ist allerdings die Gewichtung des Buches: Auf knapp 70 Seiten Erläuterungen des methodologischen Hintergrunds kommen über 100 Seiten Erläuterungen des Forschungsbeispiels, das in den Kapitel 4.5 und 4.6 etwas langatmig gerät. Dabei werden die späteren Auswertungsschritte, insbesondere die Typenbildung, sehr ausführlich am Forschungsbeispiel erläutert. Dafür überspringt Küsters die ersten Auswertungsschritte, die Anfängern meist am schwersten fallen: Sie erläutert zwar die Ergebnisse ihrer Analyse im Fallbeispiel, es bleibt aber unklar, wie sie zu diesen Ergebnissen gekommen ist.

Nohl (2006) befasst sich mit der Frage, wie man offene Interviews mit Hilfe der dokumentarischen Methode nach Ralf Bohnsack auswerten kann. Diese baut auf den zwei Sinnebenen nach Karl Mannheim auf (Kapitel 1):

Der immanente Sinngehalt bezeichnet das ‚Was‘ – das vom Interviewpartner wörtlich bzw. explizit Gemeinte. Sie lässt sich weiter

unterteilen in den intentionalen Ausdruckssinn (individuelle Handlungs- und Erzählmotive) und den unmittelbar an die Handlungspraxis geknüpften und deshalb schwer verbalisierbaren Objektsinn (soziale Bedeutung eines Textinhalts oder einer Handlung).

Der Dokumentsinn bezeichnet dagegen die Art, „wie der Text und die in ihm berichtete Handlung konstruiert ist, (...) in welchem ‚Orientierungsrahmen‘ (...) eine Problemstellung bearbeitet wird“ (S. 8). Es handelt sich hierbei um atheoretisches Wissen, d. h. um Handlungsroutinen, die Menschen verbinden, aber weitgehend unreflektiert bleiben.

Die dokumentarische Methode untersucht das Wechselspiel zwischen Was und Wie. Sie „dient der Rekonstruktion der praktischen Erfahrungen von Einzelpersonen und Gruppen, in Milieus und Organisationen, gibt Aufschluss über die Handlungsorientierungen, die sich in der jeweiligen Praxis dokumentieren, und eröffnet somit einen Zugang zur Handlungspraxis“ (S. 8). Die Auswertung besteht aus folgenden Arbeitsschritten (Kapitel 1 und 4):

Die formulierende Interpretation zielt auf den immanenten Sinngehalt, genauer: den Objektsinn ab (Der intentionale Ausdruckssinn kann mit Hilfe der dokumentarischen Methoden nicht entschlüsselt werden). Zuerst wird ein thematischer Verlauf erstellt: „Nach der Durchführung der Erhebungen hören die Forschenden die Audioaufnahmen der Interviews ab und verzeichnen tabellarisch die zeitliche Abfolge der Themen innerhalb eines jeden Falles“ (S. 46). Ausgehend von seiner Forschungsfrage wählt der Forscher diejenigen Interviewabschnitte aus, die ihn interessieren, transkribiert sie und unterzieht sie der formulierenden Feininterpretation, d. h. er identifiziert für jeden Textabschnitt Ober- und Unterthemen und fasst das thematisch Gesagte zusammen.

Die reflektierende Interpretation versucht, innerhalb eines Textes dessen dokumentarischen Sinngehalt zu rekonstruieren: Bei der formalen Interpretation mit Textsortentren-

nung wird für jeden Textabschnitt bestimmt, ob der Interviewpartner erzählt, beschreibt, argumentiert usw. Die semantische Interpretation und komparative Sequenzanalyse untersucht, wie gesellschaftliche Realität hergestellt wird, d. h. wie Äußerungen des Interviewpartners aufeinander folgen. Dies beruht auf der Annahme, dass auf „einen ersten Erzählabschnitt (...) nur ein spezifischer, nämlich ein der jeweiligen Erfahrungsweise, dem jeweiligen Rahmen entsprechender zweiter Abschnitt folgen [kann]“ (S. 11). „Forschungspraktisch wird in der Interpretation der zweite Abschnitt als gegebene, adäquate Fortsetzung eines ersten Abschnitts betrachtet und Alternativen zum zweiten Abschnitt werden gedankenexperimentell erörtert“ (S. 52). Alle denkbaren Alternativen zweiten Abschnitte (also alternative Fortsetzungen des ersten Textabschnittes) bilden zusammen den homologen Orientierungsrahmen.

Der homologe Orientierungsrahmen „wird gerade dann evident, wenn er gegenüber differentiellen Orientierungsrahmen (...) in anderen empirischen Fällen abgegrenzt werden kann“ (S. 52). Deshalb dient die komparative Analyse der Validierung von Interpretationen an anderen Textstellen.

Die komparative Analyse dient aber auch der Suche nach einem gemeinsamen Dritten und ist damit auch der erste Schritt der Typenbildung. „Die sinngenetische Typenbildung zeigt, in welchem unterschiedlichen Orientierungsrahmen die erforschten Personen jene Themen und Problemstellungen bearbeiten“ (S. 57). Dagegen „geht es in der soziogenetischen Typenbildung darum, die spezifischen Erfahrungshintergründe und die Soziogenese der Orientierungsrahmen systematisch zu analysieren“ (S. 13).

Der letzte Schritt ist die Frage, ob die entwickelte Typik generalisiert werden kann, d. h. ob sie in anderen Kontexten erhalten bleibt, auch wenn sie von anderen Typiken überlagert wird.

Da bei der qualitativen Auswertung immer

die Eigenheiten der Daten mit berücksichtigt werden, arbeitet Nohl in Kapitel 2 die Gemeinsamkeiten und Unterschiede von Experteninterviews nach Meuser und Nagel (2005), von problemzentrierten Interviews nach Witzel (1982) und von narrativen Interviews nach Schütze (1983) heraus, wobei die Darstellungen (im Gegensatz zu Küsters' Ausführungen) zu knapp sind, als dass der Leser auf dieser Basis diese Interviewformen in die Praxis umsetzen könnte. Vielmehr geht es um die methodologischen Besonderheiten, die für die spätere Interpretation wichtig sind. Zentral ist insbesondere die Unterscheidung von verschiedenen Textsorten in biographischen Interviews (Erzählungen, Beschreibungen und Argumentationen). Am Beispiel von Interviewausschnitten zeigt Nohl, woran man diese erkennt.

Um erläutern zu können, warum Interviews bevorzugt mit Hilfe der dokumentarischen Methode ausgewertet werden sollten, stellt Nohl in Kapitel 3 knapp die ersten drei Schritte der Narrationsstrukturanalyse nach Schütze dar (formale Textanalyse, strukturelle Beschreibung, analytische Abstraktion). Die Schritte 4 bis 6 (Wissensanalyse, kontrastive Vergleiche, Konstruktion eines theoretischen Modells) unterschlägt Nohl, was insofern problematisch ist, als diese, wie Küsters theoretisch und am Forschungsbeispiel gezeigt hat, „die Einzelanalysen aller Interviews eines Samples zusammenführen“ (Küsters 2006: 77). Nohl hat damit zwar Recht, wenn er an der Narrationsstrukturanalyse kritisiert: „Erst nach der intensiveren Einzelfallanalyse kommt also der Vergleich zum Zug“ (S. 37). Dennoch bleibt er den Beleg für die Behauptung schuldig, „dass es in der Narrationsstrukturanalyse zur Identifizierung des Allgemeinen und des fallspezifischen Besonderen kein methodisches Verfahren gibt, sondern das Allgemeine von dem Forschenden nur ‚vermutet‘ werden kann“ (S. 36). Eine zweite Schwäche an der Narrationsstrukturanalyse sowie an der Grounded Theory sei, dass diese soziale Phänomene eindimensional interpretierten (S. 42–43). Auch diese Kritik

greift zu kurz, da jedes qualitative Paradigma (auch die dokumentarische Methode!) eine spezifische Perspektivität einnimmt: So, wie die Biographieforschung auf Mustern in Lebensläufen abzielt, zielt die dokumentarische Methode auf die Interaktion von Handlungsrahmen und Objektsinn. Wie Küsters zeigt, können auch Lebensläufe mehrdimensional strukturiert sein (Beruf, Familie, Freizeit), und gerade der Grounded Theory geht über den Einzelfall hinaus, indem sie ein Phänomen systematisch von allen Seiten beleuchtet. Entgegen dem Eindruck, der sich bei der Lektüre von Nohls Text ergibt, ist die dokumentarische Methode damit nicht der Narrationsstrukturanalyse oder Grounded Theory überlegen, sondern es hängt von der spezifischen Forschungsfrage ab, welches Verfahren am besten geeignet ist.

Am Beispiel von Leitfaden-Interviews mit Existenzgründern über deren Haltung zu Mitarbeitern (Kapitel 5) und von narrativen Interviews zu spontanen Bildungsprozessen (Kapitel 6) präsentiert Nohl den im Rahmen der einzelnen Auswertungsschnitte produzierten Text, der Dozenten dazu dienen kann zu zeigen, wie man relevante Textstellen identifiziert und kodiert. Zum Selbststudium für Anfänger ist dieser Textteil dagegen nur bedingt geeignet, weil zu Vieles ungesagt bleibt: In Kapitel 5.1 zeigt Nohl, wie ein thematischer Verlauf eines Interviews erstellt wird und wie man aus diesem die zu transkribierenden Interviewabschnitte auswählt. Er erläutert, dass man kontrastierende Fälle zu diesen Interview aussuchen muss, erklärt aber nicht, wie man entscheidet, welche dies sind. Auch die Transkripte und Ergebnisse der formulierenden Feininterpretation werden einfach präsentiert (Kapitel 5.2 und 6.1), ohne Erläuterung dazu, was bei der Interpretation zu beachten ist. Bei der reflektierenden Interpretation (Kapitel 5.3 und 6.2) werden Textsorten mit vorher nicht definierten Begriffen benannt (etwa „Sachfrage“, „immanente Nachfrage“, „detaillierende Beschreibung“). Bei der semantischen Interpretation (Kapitel 5.3 und 6.2),

der sinngenetischen Typenbildung (Kapitel 5.4 und 6.3) und der soziogenetischen Typenbildung (Kapitel 6.4) werden die Analyseergebnisse präsentiert, aber nicht, wie man hierzu kommt, ob auch alternative Lesarten möglich gewesen wären und warum gerade diese Interpretation gewählt wurde. Für Leser, die die dokumentarische Methode nicht bereits kennen, bleibt deshalb unklar, warum die Analyse des Handlungsrahmens für die Gesamtinterpretation wichtig ist. Auch der übrige theoretisch-methodologische Rahmen ist so knapp gehalten, dass Leser, die nicht bereits mit der Ethnomethodologie, Wissenssoziologie, Biographieforschung, und Grounded Theory vertraut sind (und vor allem für solche, die aus der quantitativen Tradition kommen) Verständnisschwierigkeiten haben dürften. Hinzu kommt, dass viele zentrale Aspekte ohne Querverweise über verschiedene Kapitel verstreut sind. So erörtert Nohl in einem Unterabschnitt des Kapitels 2.2 (S. 26–28) verschiedene Textsorten, die dann später für (Kapitel 5.3 und 6.2) relevant werden.

Damit zeigt sich das Grundproblem des Buches, nämlich dass unklar ist, wer das Zielpublikum ist: Um zu vermeiden, dass sich erfahrene qualitative Forscher durch viel Bekanntes wühlen muss, wären die methodologisch interessanten und neuen Gedanken zu der Frage, „warum Interviews mit der dokumentarischen Methoden ausgewertet werden können (und sollten)“ (S. 7), besser in einem Aufsatz aufgehoben gewesen. Noch interessanter wäre ein systematischer Vergleich gewesen, wie sich die Analyseergebnisse ändern, wenn man die Texte mit der Narrationsstrukturanalyse und Grounded Theory auswertet.

Am ehesten eignet sich das Buch als Begleittext für Lehrveranstaltungen, da es reiches Anschauungsmaterial bietet und von der Verlagswebseite (www.vs-verlag.de) eine Präsentation zum Buch heruntergeladen werden kann. Doch auch hierfür bleibt das Buch unbefriedigend, da für besagte methodologischen Hintergründe andere Texte

hinzugezogen werden müssen und – was der Vergleich z. B. mit Helfferichs (2005) Einführung in die qualitative Interviewführung verdeutlicht – unklar bleibt, wie genau ein Kurs zur Interviewauswertung mit der dokumentarischen Methode aufgebaut sein könnte, also z. B. wie viele Sitzungen man zur Bearbeitung des Stoffes benötigt und welche Themen in den einzelnen Sitzungen behandelt werden sollten.

Kuckartz (2007) befasst sich „mit den neuen computergestützten Techniken des Datenmanagements und der Datenanalyse (...) Das Buch beschäftigt sich bewusst nicht mit wissenschaftstheoretischen Voraussetzungen oder methodologischen Kontroversen, sondern vertritt einen Standpunkt der Pluralität der Methoden“ (S. 9). In Kapitel 1 verdeutlicht Kuckartz, dass CAQDAS (Computer Aided Qualitative Data Analysis Software) – ähnlich wie Statistik-Software – den Forscher bei der Arbeit unterstützen, aber nicht die eigene Denkleistung ersetzen kann, und dass man auch ohne Software-Einsatz gute Forschung machen könnte, „aber es würde um einige Zehnerprozenten mehr Zeit benötigen“ (S. 13).

Kuckartz nennt kurz verschiedene Datentypen und Auswertungstraditionen qualitativer Sozialforschung. Er geht im folgenden davon aus, dass für eine spezifische Forschungsfrage Daten des gewünschten Typs – etwa Antworttexte auf offene Fragen, Interviewtranskripte, Feldnotizen, Dokumente, Texte aus Datenbanken oder Webseiten, literarische Texte, Zeitungsdaten – bereits vorliegen. In Kapitel 2 stellt Kuckartz eine Checkliste für die Textaufbereitung auf (S. 52). Insbesondere muss entschieden werden, „ob Texteinheiten definiert werden sollen (...) ob ein fixiertes Zeilenlayout erforderlich ist“ (S. 51). „Bei Texten, die noch verschriftlicht werden müssen [etwa Interviews], sind die Transkriptionsregeln zu bestimmen“ (S. 52). Anschließend müssen im QDA-Programm ein Projekt und Textgruppen eingerichtet werden und die aufbereiteten Daten importiert werden. Nun kann die eigentliche Auswertung beginnen.

Nach einem Überblick über die allen Auswertungstraditionen gemeinsamen Auswertungsschritte (Kapitel 1) erklärt Kuckartz diese im Rest des Buches Schritt für Schritt, wenn man sie mit einem QDA-Programm (Qualitativen Datenanalyse-Programm) umsetzt: „Am Anfang aller qualitativen Verfahren der Textanalyse steht eine sorgfältige Lektüre des Textmaterials“ (S. 57) und das Kodieren, d. h. das Aufbrechen des Textes und die „Zuordnung von Kategorien zu Textmerkmalen“ (S. 57).

In Kapitel 3 erläutert Kuckartz ausführlich Arten von Kategorien und des Kodierens. Typisch für die quantitative Inhaltsanalyse sind das Kodieren von Fakten und das bewertende Kodieren (Kapitel 6 und 12) mit Hilfe der lexikalische Suche, d. h. der „Möglichkeit, in den Texten oder in ausgewählten Gruppen von Texten nach dem Vorkommen bestimmter Stichworte und Wortkombinationen zu suchen“ (S. 120) und diese automatisch zu kodieren. Hierzu werden Wortlisten und Diktionäre verwendet. Bei Feldnotizen oder Dokumenten lassen sich so z. B. alle Textstellen zusammenstellen, die eine bestimmte Person betreffen. Automatisches Kodieren ist außerdem einfacher und schneller als manuelles Kodieren. I. d. R. ist aber manuelles Kodieren vorzuziehen, da dieses methodisch besser kontrollierbar ist und sich besser für die Theoriebildung eignet. In Kapitel 4 erläutert Kuckartz drei Varianten des manuellen Kodierens: theoretisches Kodieren am Beispiel der Grounded Theory (S. 72-82), thematisches Kodieren am Beispiel der Vorgehensweise der Forschergruppe um Christel Hopf (S. 83-91) sowie paraphrasierendes Kodieren am Beispiel der qualitativen Inhaltsanalyse nach Mayring (S. 91-96). Das Kapitel schließt mit einer Erläuterung der Grundprinzipien der Typenbildung (S. 96-106). Um sinnvoll mit den Kategorien arbeiten zu können, sollten nicht zu viele und nicht zu wenige existieren und sie sollten in einem linearen, hierarchischen oder netzwerkstrukturierten Kategoriensystem organisiert werden, indem Codes fusioniert oder ausdifferenziert und

Oberkategorien zu existierenden Codes hinzugefügt werden (Kapitel 11).

Sind die Daten kodiert, kann der Forscher mit Hilfe des Text-Retrievals (Kapitel 5 und 9) Text-Sets zusammenstellen lassen, die er entsprechend seinem Forschungsinteresse analysieren kann. So kann er alle Textsegmente zusammenstellen lassen, die einem bestimmten Code zugeordnet sind (Einfaches Retrieval, S. 110-111), die gleichzeitigen mehreren Codes zugewiesen wurden (Kontrastierendes Retrieval, S. 111-112) oder in denen sich mehrere Codes überschneiden (Verknüpfendes Retrieval, S. 112-113). Er kann Tabellen über die Häufigkeit von Codkombinationen erstellen lassen (S. 115-117; 157-159), Subgruppen von Daten vergleichen (S. 155-157), mit UND-, ODER-, Sequenz-, Entfernungs-, Überschneidungs- und Einbettungs- und Mengenoperatoren nach Mustern von Kodierungen suchen (S. 159-164) und Hypothesen überprüfen (S. 164-171).

Ein wesentliches Instrument qualitativer Sozialforschung sind Memos, also Notizen des Forschers etwa zur Datenerhebungssituation, zur inhaltlichen Bedeutung von Codes, Zusammenfassungen des Inhalts von Interviews, Hypothesen oder theoretische Überlegungen (Kapitel 7).

Wenn ein Forscher qualitative und quantitative Sozialforschung kombinieren will, aber auch zur Vereinfachung des Text-Retrievals, bietet sich die Möglichkeit an, Fallvariablen anzulegen, d. h. nicht Textstellen, sondern einem ganzen Text Ausprägungen auf einer Variable zuzuweisen. Diese können entweder aus Rahmendaten zum Text stammen oder Ergebnisse der Datenanalyse sein (Kapitel 8). Exportiert man diese Variablen z. B. in SPSS, kann man sie statistisch weiterbearbeiten, wie Kuckartz in Kapitel 13 am Beispiel der Cluster- und Faktorenanalyse illustriert.

Neu gegenüber der vorigen Auflage ist nicht nur ein Index, sondern auch ein eigenes Kapitel über Daten-Display und Visualisierung: In Kapitel 10 erläutert Kuckartz, dass

QDA-Software selbst mit Symbolen und Farben arbeitet, um die Auswertungsarbeit zu erleichtern. Weiterhin können Analyseergebnisse selbst visualisiert werden, indem „Daten, Themen, Kategorien oder Konzepte in Bilder übersetzt“ (S. 180) werden. Beispiele sind fallbezogene Visualisierungen, fallübergreifende Visualisierungen und Konzept Maps.

Das Buch schließt mit einem Ausblick über die Zukunft der computergestützten qualitativen Datenanalyse, einem Überblick über gängige CAQDAS sowie Entscheidungshilfen für die Wahl der richtigen Software.

Wegen seiner starken Praxisorientierung eignet sich die Einführung sowohl zum Selbststudium, als auch zum Einsatz in Lehrveranstaltungen: Kuckartz führt in jedem Kapitel zunächst allgemein und anhand von Forschungsbeispielen in die Problematik ein, nennt die zu treffenden Entscheidungen samt Implikationen und erläutert die programmtechnische Umsetzung am Beispiel des von ihm entwickelten Programms MAXqda. Das Kapitel schließt mit Übungsaufgaben, zu denen ein Tutorial auf der Webseite (www.maxqda.de) existiert. Das Buch eignet sich damit sowohl als generelle Einführung in qualitative Auswertungstechniken, als auch in CAQDAS, sollte aber sowohl um einen Text zur Datenerhebung des gewünschten Datentyps, als auch um einen Text zur gewählten Auswertungstradition – etwa Strauss und Corbin (1996) oder Mayring (2003) – ergänzt werden. Auch wenn die Anwendungsbeispiele sich auf das von Kuckartz entwickelte QDA-Programm MAXqda beziehen, lassen sie sich problemlos auf andere CAQDAS übertragen, v.a. weil.

Alle drei Texte sind gute Beispiele dafür, dass qualitative Sozialforschung – entgegen verbreiteter Vorurteile – sehr wohl lehrbar ist, zeigen aber auch die Probleme auf:

Zunächst wird sehr deutlich, dass die Tradition qualitativer Lehrbücher im deutschen Sprachraum noch sehr jung ist, so dass viele Fragen noch ungeklärt sind und die Bücher

deutlich verbesserungswürdig sind, was sich u. a. auch darin zeigt, dass Kuckartz' Buch, das ja mittlerweile in der dritten Auflage erschienen ist, deutlich am ausgereiftesten ist. Darüber hinaus ergeben sich für Lehrbuchautoren Schwierigkeiten, die in der qualitativen Sozialforschung selbst begründet sind und auch in Zukunft das Verfassen von einzelnen Lehrbüchern erschweren und einen so einheitlichen Methodenkanon wie in der quantitativen Sozialforschung verhindern könnten: Zum einen liegt ja gerade die Stärke qualitativer Sozialforschung darin, dass sehr verschiedene theoretische Fragen an sehr unterschiedliche Datentypen gestellt werden und damit der Forschungsprozess auch sehr flexibel ist. Zum anderen lassen sich Datenerhebung und -auswertung zwar analytisch, nicht aber forschungspraktisch trennen. Deshalb hat qualitative Forschung mit quantitativer Forschung gemein, dass Lehrbücher zwar Grundlagen legen können, viele Probleme und Kniffe dann aber doch in der Praxis – am besten in einem Forschungsprojekt – gelernt werden müssen. Ähnlich wie man für die quantitative Sozialforschung zumindest eine Einführung in die Datenauswahl und -erhebung und eine in die Statistik benötigt, genügt – das zeigen die hier besprochenen Bücher – auch für die qualitative Sozialforschung kein einzelnes Methodenbuch.

Literatur

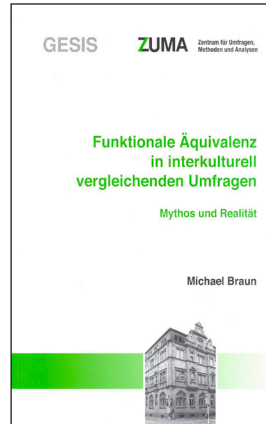
- Fuchs-Heinritz, W., 2005: Biographische Forschung. Wiesbaden: VS-Verlag.
- Helfferich, C., 2005: Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews. Wiesbaden: VS-Verlag.
- Mayring, P., 2003: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim/Basel: Beltz.
- Meuser, M. und U. Nagel, 2005: ExpertInneninterviews – vielfach erprobt, wenig bedacht. S. 71-94 in: Bogner, Alexander et al. (Hg.): Das Experteninterview. Opladen: VS-Verlag.
- Schütze, F., 1983: Biographieforschung und narratives Interview. Neue Praxis 13: 283-293.

Strauss, A. und J. Corbin, 1996: Grundlagen qualitativer Sozialforschung. Weinheim: Beltz

Witzel, A., 1982: Verfahren der qualitativen Sozialforschung – Überblick und Alternativen. Frankfurt am Main. New York: Campus.

NINA BAUR, BERLIN

* * * * *



MICHAEL BRAUN, 2006.
Funktionale Äquivalenz
in interkulturell vergleichenden Umfragen.
Mythos und Realität.
GESIS – ZUMA.
ISBN 978-3-924220-33-4,
294 Seiten, 13 EUR.

Interkulturell vergleichende Forschungsprogramme wie der World Value Survey oder das International Social Value Programme stellen seit einigen Jahrzehnten eine wichtige Grundlage für den soziologischen Gesellschaftsvergleich dar. Studien dieser Art bieten nicht nur die Möglichkeit, den Wandel von sozialen Verhaltensweisen und Wertorientierungen auf globaler Ebene systematisch zu dokumentieren und zu beschreiben, sondern auch die sozialstrukturellen und kulturellen Bedingungen, die diesem Wandel zugrunde liegen, theoriegeleitet zu untersuchen. Jeder, der sich selbst ausführlicher mit dieser Form von Forschung beschäftigt, weiß allerdings, dass die Konstruktion funktional äquivalenter Erhebungsinstrumente, die über Länder und Kulturkreise hinweg ein valides Abbild der anvisierten Verhaltens- oder Einstellungsdimension liefern, eine schwierige Aufgabe darstellt und dass dementsprechend bei der Interpretation der Ergebnisse große Vorsicht geboten ist.

Michael Braun hat sich in seiner Habilitationsschrift die Aufgabe gestellt, diese Problematik sowohl theoretisch als auch empirisch zu untersuchen. Im ersten Teil der Arbeit befasst er sich mit theoretischen Ansätzen zur Erklärung des Befragtenverhaltens, mit möglichen Fehlerquellen in der vergleichenden Umfrageforschung sowie mit methodologischen Strategien zur Überprüfung der Äquivalenz von Fragebogenitems. Die interkulturelle Vergleichbarkeit von Umfragedaten kann dadurch beeinträchtigt werden, dass bei der Übersetzung eines Items in verschiedene Sprachen aufgrund kulturspezifischer Kommunikationsnormen leichte Veränderungen der Frage- und/oder Antwortformulierung erforderlich sind, wodurch sich der Bedeutungsgehalt ändert. Wie Braun nachweist, passieren zudem mitunter auch ganz offensichtliche Übersetzungsfehler. Noch wichtiger als diese Art von Fehlerquellen sind jedoch nach Braun sogenannte Kontext- und Framingeffekte. In Anlehnung an Stocké unterscheidet er drei Typen von Effekten (S. 58f.): Ambiguitätsbasierte Framingeffekte kommen dadurch zustande, dass bestimmte Komponenten der Frageformulierung nicht ausreichend erläutert werden, wodurch der Befragte die Möglichkeit hat, die Frage in Abhängigkeit von seinem persönlichen und kulturellen Hintergrundwissen unterschiedlich zu interpretieren. Heuristikbasierte Framingeffekte liegen dann vor, wenn der Befragte seine Antwort nicht auf den gesamten Text des Items, sondern nur auf einen Teilaspekt bezieht. Schemabasierte Framingeffekte schließlich beruhen darauf, dass durch bestimmte Formulierungen gruppen- oder kulturspezifische mentale Schemata – wie etwa Ideologieschemata – aktiviert werden, die das Antwortverhalten mitbestimmen; die Antworten beziehen sich in diesem Fall nicht nur auf den expliziten Fragetext, sondern auf das kognitive Schema, das beim Befragten aktiviert wurde. In diesem Zusammenhang ist auch die Frage von Bedeutung, ob bestimmte Items von den Befragten als Meinungs- oder als Einstel-

lungsitems interpretiert werden, das heißt ob sie mit ihrer Antwort eine Einschätzung der allgemeinen Lage in ihrem sozialen Umfeld (Meinung) oder ihre persönliche Wertorientierung (Einstellung) zum Ausdruck bringen. Dieses Problem stellt sich nicht zuletzt deshalb, weil in der Umfrageforschung die Technik der indirekten Ermittlung von Einstellungen mittels Meinungsfragen vielfach ganz bewusst eingesetzt wird.

Auf der Basis dieser Vorüberlegungen untersucht Braun im zweiten Teil seiner Arbeit die funktionale Äquivalenz von Umfragedaten am Beispiel der Einstellungen zu Geschlechtsrollen. In einem ersten Schritt werden die entsprechenden Items des International Social Survey Programme zum Thema Family and Gender Roles einer Reanalyse unterzogen. Die Analysestrategie besteht darin, durch die Überprüfung der internen Konsistenz von Skalen wie auch durch die Untersuchung der externen Konsistenz einzelner Items bezüglich relevanter Außenkriterien (zum Beispiel Alter) Anhaltspunkte für länderspezifische Textinterpretationen und Antworteffekte zu gewinnen. So wird beispielsweise aus dem Vorhandensein einer niedrigen Korrelation zwischen dem Item „Hausfrau zu sein ist genauso erfüllend, wie gegen Bezahlung zu arbeiten“ und der Variable Alter der Schluss gezogen, dass diese Frage im entsprechenden Land eher als Meinungsitem interpretiert wurde; eine hohe Korrelation deutet hingegen nach Auffassung Brauns darauf hin, dass die Geschlechtsrollenthematik in diesem Land viel mehr nach ideologischen Gesichtspunkten beurteilt wird, die stark nach Alterskohorten variieren. Mit einer Vielzahl von Detailanalysen wird ein deutlicher Nachweis erbracht, dass bei einer erheblichen Zahl der ISSP-Items starke länderspezifische Kontext- und Framingeffekte auftreten, wodurch eine Interpretation der Ergebnisse im Sinn der intendierten Dimension nur in eingeschränktem Maße möglich ist.

Im zweiten Schritt seiner Untersuchungen versucht Braun, bestimmte Kontexteffekte,

die in den ISSP-Daten festgestellt werden konnten, durch die Analyse einer interkulturellen Vergleichsstudie (Thyssen-Studie), in der zusätzlich zu den ISSP-Fragen eine Reihe weiterer Items zum Thema Geschlechtsrollen erhoben wurden, näher zu beleuchten. Da auch hier zahlreiche Detailanalysen vorgenommen wurden, soll die Methodik wiederum an einem konkreten Beispiel skizziert werden: Um zu überprüfen, ob die länderspezifische Bedeutung der Erwerbsarbeit die Beantwortung des oben genannten Items „Hausfrau zu sein ist genauso erfüllend, wie gegen Bezahlung zu arbeiten“ beeinflusst und somit die Brauchbarkeit dieses Items als Indikator für die Geschlechtsrollen-Ideologie beeinträchtigt, wurde in die Thyssen-Studie ein zusätzliches Item eingeschaltet, in dem die Befragten die entsprechende Tätigkeit eines Hausmanns beurteilen sollten. Wenn man sich nur auf die Ergebnisse des ersten Items („Hausfrau zu sein ist genauso erfüllend ...“) stützt, müsste man annehmen, dass in Ostdeutschland egalitäre Geschlechtsrollen-Einstellungen weiter verbreitet sind als in allen anderen untersuchten Ländern, da hier die Hausfrauenrolle am stärksten abgelehnt wird. Aus der gemeinsamen Betrachtung der Ergebnisse beider Fragen (Hausfrau und Hausmann) wird jedoch ersichtlich, dass Ostdeutsche die Tätigkeit eines Hausmanns noch erheblich negativer beurteilen als die Tätigkeit einer Hausfrau, während etwa Kanadier und Spanier diesbezüglich kaum Unterschiede machen. Die differenzierte Analyse macht somit deutlich, dass die Ablehnung der Hausfrauenrolle in Ostdeutschland nicht nur auf egalitäre Geschlechtsrollenorientierungen, sondern auch auf den hohen ideologischen Stellenwert und die finanzielle Notwendigkeit der Erwerbsarbeit für beide Geschlechter zurückzuführen ist.

Die Problematik der Interpretation von Umfrageitems wird schließlich auch anhand einer deutschen multipurpose-Studie dokumentiert, in der Befragte um eine offene Begründung für die von ihnen vergebenen Skalenwerte gebeten wurden. Diese Unter-

suchung bestätigt nachdrücklich, dass der Beantwortung derartiger Fragebogenitems unterschiedliche Überlegungen zugrunde liegen können. So wählten beispielsweise manche weibliche Befragte beim Item „Hausfrau zu sein ist genauso erfüllend ...“ als Antwort die neutrale Mittelkategorie, weil sie glauben, dass manche Frauen die Hausfrauenrolle durchaus als befriedigend erleben, obwohl diese Rolle für sie selbst absolut unerfüllend wäre, während andere Personen mit der selben Einstellung dieses Item klar ablehnten. Umgekehrt kann das Vorhandensein ideologischer Schemata dazu führen, dass Befragte mit ähnlichen Einstellungen zu ein und demselben Item sogar konträre Antworten geben – so etwa, wenn Personen, die eine egalitäre Geschlechterideologie vertreten, den Rollentausch („Der Mann sollte zu Hause bleiben und sich um die Kinder kümmern, während die Frau arbeitet“) in einem Fall befürworten, im anderen Fall aber ablehnen, weil sie der Meinung sind, dass die Haus- und Familienarbeit für beide Geschlechter gleichermaßen unerfüllend ist.

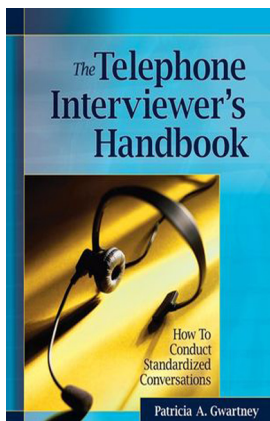
Ein grundsätzliches Problem der ISSP-Items zum Thema Geschlechtsrollen liegt nach Braun darin, dass den verwendeten Messinstrumenten das traditionelle Geschlechtsrollenmodell der 1960er Jahre zugrunde liegt. Die Replikation der altbewährten Items ermöglicht die Generierung von exakt vergleichbaren Zeitreihendaten. Dieser methodologische Gewinn geht aber zulasten der substantziellen Aussagekraft, da sich die alten Messinstrumente nicht dazu eignen, neue Trends und Geschlechtsrollenbilder zu erfassen. Wie jedoch Brauns Analysen zeigen, ist auch die Aufnahme neuer Items, die aktuelle Entwicklungstrends (zum Beispiel die Frage des Rollentauschs) thematisieren, problematisch, da derartige Items von den Befragten in traditionelleren Ländern vielfach nicht verstanden werden oder andere Assoziationen erwecken als in ökonomisch hochentwickelten westlichen Gesellschaften.

Michael Braun hat mit diesem Buch einen insgesamt sehr gelungenen Beitrag zur Me-

thodik und Problematik der interkulturell vergleichenden Umfrageforschung vorgelegt. Seine empirischen Untersuchungen belegen in exemplarischer Weise, welche Fallstricke diese Art von Forschung beinhaltet und an welche methodologischen Grenzen man stößt, wenn man eine Vielzahl sehr heterogener Gesellschaften anhand derselben Fragebogenitems vergleichen will. Zugleich bestätigen sie aber auch, dass diese Forschung sehr gewinnbringende soziologische Gesellschaftsvergleiche ermöglicht, wenn der Forscher mit der Gesellschaft (Sozialstruktur, Kultur, Sprache) der von ihm untersuchten Länder in ausreichender Weise vertraut ist und möglicherweise auftretende methodologische Verzerrungseffekte bei der Analyse und Interpretation der Daten in entsprechender Weise berücksichtigt.

FRANZ HÖLLINGER, GRAZ

* * * * *



PATRICIA A. GWARTNEY, 2007. *The Telephone Interviewer's Handbook. How To Conduct Standardized Conversations*. San Francisco: Jossey-Bass. ISBN 978-0-7879-8638-4, 336 Seiten, 22,90 EUR.

Die Bedeutung des Interviewerverhaltens für die Datenqualität wird in allen Standardtexten über Methoden der empirischen Sozialforschung und der Umfrageforschung angesprochen. Interviewer sind als Umfrageoperatoren verantwortlich für Designeffekte und in der Kontakt- und Interviewphase mitbestimmend für das Ausmaß von Unit- und Item-Nonresponse. Sie stellen ferner wichtige Kodeterminanten des Messfehlers. Angesichts der Zentralität des In-

terviewerverhaltens für Stichproben- und Messqualität muss erstaunen, dass die Anzahl der Forschungsarbeiten zum Interviewerverhalten im Vergleich zur Anzahl der Arbeiten über andere Determinanten der Datenqualität eher gering ausfällt. Im Vergleich zur Literatur über das Testen und die Evaluation von Erhebungsinstrumenten sowie Pretestverfahren, in der durchaus auch praktische Empfehlungen zur optimalen Gestaltung von Erhebungsinstrumenten zu finden sind, fehlen in Hinblick auf das Interviewerverhalten zusammenfassende, die verfügbare wissenschaftliche Literatur einbeziehende, detailliertere praktische Empfehlungen zur Optimierung des Interviewerverhaltens, sieht man von verstreuten, z.Teil dem Leser zur eigenen Ableitung aus den wissenschaftlichen Studien überlassenen Empfehlungen in den Textbüchern und in einzelnen Aufsätzen ab. Die Schulungsregeln zum Training von Interviewern gehören zumeist zum Insiderwissen der Institute und erblicken das Licht einer beschränkten Öffentlichkeit erst, wenn der Auftraggeber dies fordert. Die im Internet zu findenden Empfehlungen sind vollkommen unzureichend.

In diesem Zusammenhang versucht die als Handbuch für Telefoninterviewer konzipierte Veröffentlichung von Patricia Gwartney (Professor of Sociology, University of Oregon) eine wichtige Lücke im Bereich der computerunterstützten Telefonumfragen zu schließen. Die Autorin gründete im Jahr 1992 das Oregon Survey Research Laboratory, welches sie bis 2003 leitete und in dem sie ca. 290 verschiedene Studien betreut hat. In ihrem Handbuch wendet sich Patricia Gwartney nicht an Forscher oder Trainingsleiter, sondern direkt an die Interviewer: „This handbook speaks directly to you, new interviewers just starting your training. It aims to teach you basic principles and procedures for conducting telephone interviews, from the moment you dial the telephone to the moment you say good-bye“ (S. 1). In dieser direkten und sehr einnehmenden Sprache vermittelt die Autorin nicht nur die wichtigen Prin-

zipien der Durchführung standardisierter Telefoninterviews, sondern sehr viel mehr. Die Interviewer lernen Typen von Umfragenorganisationen kennen, sie erfahren, was sie selber über ihre Arbeitgeber wissen sollten und welche Fragen sie ihnen stellen sollten. Sie erfahren ferner etwas über ethische Richtlinien, über das, was sie in der Interviewerschulung erwartet, über die für ihre praktische Arbeit wichtigen technischen Aspekte des CATI-Verwaltungsprogramms wie Anruf- und Nummernverwaltung, Dispositionscodes etc., über das optimale Verhalten in standardisierten Interviews, über Formen der Kontaktaufnahme und der Auswahl von Personen im Haushalt, über den Umgang mit weichen und harten Verweigerern einschließlich der Methoden des „refusal conversion“ und über die Kontrolle der eigenen Stimme (voice management) allgemein, aber auch im Umgang mit bestimmten Typen von Verweigerern. Detailliert wird aufgeklärt, wie Fragen gestellt, wie man mit sensiblen Fragen umgeht und durch welche Arten von Probing inadäquate Antworten geklärt werden sollten. Schließlich wird den Interviewern auch vermittelt, wie ihr Arbeitsplatz eingerichtet ein sollte und wie die Routinen am Arbeitsplatz bei der alltäglichen Arbeit aussehen. Detailliert werden auch wichtige ergonomische Hinweise gegeben.

Das Buch ist außerordentlich anschaulich und praxisnah geschrieben. In zahlreichen Übersichten sind Probleme und Beispiele aufgelistet. Die Interviewer erfahren neben den Problemen stets auch, wie sie diese Probleme durch Modifikation des eigenen Verhaltens lösen können und welche Konsequenzen ein mögliches Fehlverhalten hätte. Es finden sich viele Beispiele für die Kontaktaufnahme und die Aufrechterhaltung von Kontakten sowie für das Screening. Man merkt dem Buch nicht nur an, dass es von einer langjährig erfahrenen Praktikerin geschrieben wurde, sondern dass es auf einem soliden Studium der einschlägigen Literatur beruht. Es enthält nicht nur ein Glossar der wichtigsten einschlägigen Begriffe sondern

auch ein Literaturverzeichnis, so dass interessierte und entsprechend vorgebildete Interviewer den Stoff weiter vertiefen können. Vor allem aber ist das Buch interessant, fesselnd und in einer für die Interviewer verständlichen Sprache geschrieben. Es enthält das für die Interviewer wirklich notwendige Wissen. Das Buch sollte nicht nur jedem, der eine Interviewertätigkeit aufnimmt, sondern auch jedem Interviewer, der in einem methodisch anspruchsvollen Institut tätig ist, an die Hand gegeben werden. Es ist nicht nur geeignet als Handbuch für Interviewer, sondern stellt darüber hinaus auch einen Schulungsleitfaden für Schulungsverantwortliche dar und trainiert insofern auch die Trainer. Werden die in dem Buch gegebenen Empfehlungen real umgesetzt, wird dies der Datenqualität mit Sicherheit zugute kommen. Es wäre dringend zu empfehlen, das Buch zu übersetzen und deutschen Interviewern zugänglich zu machen. Schließlich ist das Buch auch in Lehrveranstaltungen über Methoden und Praxis der Umfrageforschung einsetzbar.

FRANK FAULBAUM, DUISBURG

* * * * *



MICHAEL HÄDER, 2006.
Empirische Sozialforschung. Eine Einführung.
Wiesbaden: VS Verlag.
ISBN 978-3-531-14010-0,
497 Seiten, 19,90 EUR.

Ein einfaches Projekt ist es gewiss nicht, ein neues Lehrbuch zu den Methoden der empirischen Sozialforschung vorzulegen, gerade wenn einige gute Lehrbücher (auch in Neuauflagen) zur Verfügung stehen. Mit dem Vorwurf „alten Wein in neuen Schläuchen“

wiederzugeben, ist man schnell konfrontiert. Innovativ ist ein Methodenlehrbuch aber genau dann, wenn es neuere methodologische Entwicklungen aufgreift und diese in die klassischen Themengebiete integriert. Diese Vorgehensweise hat Michael Häder mit seinem Buch recht konsequent umgesetzt.

Zunächst beschreibt er in Kapitel 1 kurz den Aufbau des Buches, während Kapitel 2 sich beispielorientiert mit der Notwendigkeit von Methodenwissen in verschiedensten empirisch orientierten Disziplinen auseinandersetzt und Grundbegriffe definiert. Kapitel 3 setzt sich mit den wissenschaftstheoretischen Grundlagen der empirischen Sozialforschung auseinander. Neben den üblichen Darstellungen über die Bedeutung von Theorien, Hypothesen, Operationalisierungen und dem Werturteilsproblem wird auch ausführlicher auf die unterschiedlichen Methodologien qualitativer und quantitativer Forschungsansätze eingegangen. Kapitel 4 befasst sich mit dem Forschungsprozess, Mess- und Skalierungsverfahren sowie Erhebungsdesigns. Bei letzteren geht Häder erfreulicherweise ausführlicher auf verschiedene Arten von Paneldesigns ein. Allerdings fehlt die Darstellung der Kombination von Panel- und Kohortendesign, die insbesondere bei Längsschnituntersuchungen von Schülerpopulationen eingesetzt wird. In Kapitel 5 werden Stichprobenverfahren behandelt und auch neuere Entwicklungen (insbesondere Stichproben für Access-Panel und Internetbefragungen) berücksichtigt. Sehr anschaulich ist auch die Erörterung des ADM-Stichprobendesigns einschließlich der notwendigen Schritte, um zu einer brauchbaren Zufallsstichprobe zu gelangen. Hier wird Studierenden, die sich im Hinblick auf Tätigkeiten in der Markt- und Meinungsforschung orientieren wollen, ein solides Grundwissen vermittelt. Ein weiteres, in anderen Lehrbüchern nur sporadisch behandeltes Thema bezieht sich auf Stichproben für interkulturelle Studien, wobei Häder insbesondere auf die Studien des European Social Survey (ESS) eingeht.

Die Erhebungsmethoden (Kapitel 6) gliedern sich klassisch nach Befragung, Beobachtung und Inhaltsanalyse. Die verschiedenen Befragungstechniken, die in der methodologischen Literatur diskutierten „Theorien der Befragung“ sowie die Verzerrungsaspekte machen den größten Teil des Kapitels aus. Erfreulich ist die Übersicht auf Seite 283, die die verschiedenen neueren, insbesondere auf Computerunterstützung basierenden Varianten der schriftlichen, telefonischen und mündlichen Befragungstechniken enthält. Ebenso werden dem Leser hier die in der Lehre oft verwendeten ALLBUS-Studien, der schon genannte ESS, der Mikrozensus und das sozioökonomische Panel (SOEP) vorgestellt. Aufgrund der vielfach höheren Anwendungshäufigkeit von Befragungen in der Sozialforschung ist ein stärkeres Gewicht in der Darstellung im Vergleich zu Beobachtungstechniken und inhaltsanalytischen Verfahren in diesem Kapitel durchaus gerechtfertigt. Unter dem Kapitel „Komplexe Designs“ (Kapitel 7) werden Experimente, Fallstudien, Delphi-Befragungen, Evaluationsstudien, interkulturelle Studien sowie Zeitbudgetstudien behandelt. Bis auf die Experimente und die Fallstudien werden die anderen Themen nur relativ kurz (oder gar nicht) in anderen Lehrbüchern zur empirischen Sozialforschung erwähnt. Der Delphi-Ansatz wird bisher nur in der speziellen Fachliteratur behandelt. Die deutliche Zunahme von Evaluationsstudien rechtfertigt, dass Studierende früh mit dem Design vertraut gemacht werden. Die Komplexität interkultureller Studien am Beispiel des ESS weist auf die Spezifika ländervergleichender Erhebungen hin. Neben den Datensätzen des ALLBUS werden auch die des ESS in Zukunft verstärkt in der Lehre eingesetzt werden, was durch die einfache Zugänglichkeit über die entsprechenden Webseiten gefördert wird. Auch hier ist das Lehrbuch von Häder eine hilfreiche Grundlage.

Ungewöhnlich für ein Methodenlehrbuch ist ein eigenständiges Kapitel zu „Pretests“ (Kapitel 8). Dies ist insofern gerechtfertigt,

als sich neben der steigenden Bedeutung verschiedener Varianten von Erhebungstechniken auch die Pretestverfahren weiter entwickelt haben. Häder differenziert neben den im Feld verwendeten Pretests die kognitiven Verfahren – die in den letzten Jahren besondere Bedeutung im Kontext der ZUMA-Methodenforschung erhielten – sowie Fokusgruppen bzw. Expertenbewertungen. Am Ende des Kapitels wird anhand einer empirischen Untersuchung im Forschungskontext des Autors (Dresdner Notwehrbefragung) ein kognitives Pretestdesign schrittweise und sehr anschaulich erläutert.

Datenaufbreitung und -auswertung ist Gegenstand des 9. Kapitels. Es werden zunächst Beispiele zur Aufbereitung qualitativer und quantitativer Daten gegeben. In den „Grundprinzipien der statistischen Analysen“ wird relativ knapp auf bekannte deskriptive und inferenzstatistische Auswertungstechniken eingegangen. Multivariate Techniken beschränken sich auf ein exploratives Verfahren (CHAID-Analyse), die Facettentheorie und die multidimensionale Skalierung (MDS). Letzteres wird beispielhaft über die schon erwähnte Dresdner Studie erläutert. Die Nutzung der Facettentheorie zur Datenauswertung und die Erörterung verschiedener MDS-Lösungen können beim statistisch weniger versierten Leser den Eindruck hinterlassen, als wenn MDS ein sehr verbreitetes Verfahren für Umfragedaten wäre. Hier wird das Lehrbuch eindeutig zu speziell. Häufig verwendete explorative statistische Verfahren zu Item- und Skalenanalysen, wie beispielsweise die Faktorenanalyse, werden ebensowenig erwähnt, wie hypothesentestende, konfirmatorische Techniken (beispielsweise die multiple Regressionsanalyse und die Pfadanalyse).

Statistische Auswertungstechniken können in einem Methodenlehrbuch sicherlich keinen großen Raum einnehmen. Daher wird auch von Häder an verschiedenen Stellen auf geeignete Statistiklehrbücher hingewiesen. Leider beschränkt sich die Erörterung qualitativer Auswertungstechniken gera-

de einmal auf eineinhalb Seiten, wobei im Unterschied zu den Verweisen auf die Statistiklehrbücher hier kein Standardwerk zu qualitativen Auswertungsverfahren angegeben wird. Ein Verweis auf das Lehrbuch von Siegfried Lamnek hätte hier gut berücksichtigt werden können. Mit der „Dokumentation empirischer Projekte“ (Kapitel 10) wird das Lehrbuch abgeschlossen. Häder geht hier auf Qualitätskriterien der Umfrageforschung und die Notwendigkeit von Methodenberichten bei größeren und längerfristigen Erhebungen ein.

Insgesamt betrachtet kann das Buch als eine Bereicherung für die Lehre in empirischer Sozialforschung angesehen werden. Seine Stärken liegen zweifelsfrei in den Kapiteln zu den Auswahl- und Erhebungsverfahren. Hier werden neueste Techniken erörtert, die Literaturverweise und die Internetquellen sind jeweils recht aktuell. Die Wissenschaftstheorie kommt eher etwas zu kurz und das Kapitel zur Datenanalyse eignet sich nicht, einen Überblick über die gängigen quantitativen und qualitativen Auswertungstechniken zu erhalten. Hier sollte man stattdessen direkt auf die jeweiligen Lehrbücher zurückgreifen.

JOST REINECKE, BIELEFELD

Ankündigungen

G.O.R. 08
10th International G.O.R Conference
GENERAL ONLINE RESEARCH

Universität Hamburg
10. -12. März 2008

G.O.R. 08
10th International
G.O.R Conference
GENERAL ONLINE
RESEARCH

Vom 10. -12. März 2008 findet an der Universität Hamburg die 10. Internationale Konferenz GENERAL ONLINE RESEARCH statt. Die Tagung wird von der Deutschen Gesellschaft für Onlineforschung e.V. (DGOF e.V.) zusammen mit EARSandEYES GmbH sowie Prof. Sattler, Lehrstuhl für Marketing, organisiert. In über 100 wissenschaftlichen Beiträgen aus dem In- und Ausland werden Theorien, Methoden, Techniken und Ergebnisse der Forschung über das Internet sowie der Online- oder Mobilkommunikation behandelt. Ziel der G.O.R 08 ist es, den Stand der wissenschaftlichen Forschung, innovative Entwicklungen sowie praktische Erfahrungen zu dokumentieren und damit einen Transfer zwischen

- Wissenschaftlern
- Anwendern der Internet-Forschung
- Theorie und Praxis
- Universitäten und Unternehmen
- Nachfragern und Anbietern

zu ermöglichen. Eine detaillierte Auflistung aller Themenbereiche sowie die Anmeldemodalitäten sind der folgenden Website zu entnehmen: <http://www.dgof.de/>

Workshop

Beginning to Advanced Methods in Structural Equation Modeling

Peter M. Bentler

University of California at Los Angeles

University of Duisburg-Essen

February 10 - 13, 2008

Advanced Methods in Structural Equation Modeling

This 2.5 day workshop covers beginning to advanced topics in structural equation modeling (SEM). Each participant will get a free (time-limited) copy of the EQS program for use on their computer. A background in basic statistics and linear regression is required; exposure to matrix algebra will be helpful but not necessary.

Practical issues (modeling with Diagrammer or text files) and theoretical topics (use of recently published and in press statistics) will be covered, depending on participant interest:

Typical models. Latent variables. Factor scores. Identification, estimation, testing, interpretation.

Model modification: z, LM, and Wald tests. Fit indices and power. Mean structure models. Growth curve models, and their automatic setup.

Best practices with nonnormal data: Satorra-Bentler tests and Yuan-Bentler residual-based statistics. Missing data estimators, tests, and SE's for normal (Jamshidian-Bentler) and nonnormal (Yuan-Bentler) data. Kim-Bentler tests of homogeneity of means and covariance matrices. Correlation structure methods for normal and nonnormal data, and categorical variables. Simulation, bootstrap and jackknife methods. Case-weighting for complex samples and robust statistics (Yuan-Bentler) for outliers and influential points. Multi-sample/multi-group models. Multilevel models with ML (Bentler-Liang) and robust tests and SEs (Yuan-Bentler); HLM approaches. Mixture models. Internal consistency reliabilities: alpha, 1-factor, x-factor, arbitrary model, maximal, dimension-free, greatest lower bound.

Peter M. Bentler has been an elected president of the Society of Multivariate Experimental Psychology, the Psychometric Society, and the Division of Evaluation, Measurement, and Statistics of the American Psychological Association. He is the elected 2007-8 President of the Western Psychological Association. Karl Jöreskog and he are the joint 2007 recipients of the American Psychological Association's Distingu-

ished Scientific Contribution Award for the Applications of Psychology.

Registration and further information:

<http://www.uni-due.de/soziologie/workshop.eqs/index.shtml>

Location: *University of Duisburg-Essen, Mercator-Haus, Forsthausweg, Übungsraum LF 052, 47048 Duisburg*

* * * *

Workshop

Einführung in familien- und haushaltsbezogene Analysen mit dem Mikrozensus

GESIS-ZUMA, Mannheim

11. - 12. März 2008

Familien- und haushaltsbezogene Analysen mit dem Mikrozensus

Der Mikrozensus bietet aufgrund seiner Anlage als Haushaltsstichprobe – befragt werden alle in einem Haushalt wohnenden Personen – ein wichtiges Datenpotenzial für haushalts- und familienbezogene Analysen mit sozialstrukturellem Schwerpunkt. Ziel dieses Workshops ist es, einen Einblick in dieses Potenzial zu vermitteln, aber auch vorhandene Grenzen aufzuzeigen. Hierfür wird am ersten Tag ein Überblick in die spezifischen Besonderheiten des Mikrozensus mit dem Schwerpunkt ‚Haushalt und Familie‘ gegeben. Am zweiten Tag erfolgt eine Vertiefung dieser Kenntnisse anhand von praktischen Übungen (mit SPSS) und konkreten Analysebeispielen.

Der Workshop wird in enger Kooperation mit dem Statistischen Bundesamt (Gruppe VIII C – Mikrozensus, Arbeitskräftestichprobe, Haushalt und Familie) durchgeführt.

Interessenten werden gebeten, sich beim

GESIS-ZUMA-Tagungssekretariat anzumelden:

workshop-mannheim@gesis.org; Tel.: 0621-1246-221

Teilnahmebeitrag: 120 €, Studenten 80 €

Kontakt:

Dr. Heike Wirth

GESIS-ZUMA

Postfach 12 21 55

68072 Mannheim

Tel.: 0621-1246-269 FAX 0621-1246-100

heike.wirth@gesis.org

<http://www.gesis.org/Dauerbeobachtung/GML/index.htm>

Call for Papers

Interims Meeting of the ISA (International Sociological Association) Research Committee 33 (RC33) on „Logic and Methodology in Sociology“

Naples, Italy

September 1 - 5, 2008

Logic and Methodology in Sociology

In addition to great pizza, pasta, and wine, we are happy to announce that we will have approximately 100 sessions covering all topics in quantitative and qualitative methods of empirical social research, including applied statistics. RC33 is interdisciplinary; as it holds for the last conferences in Cologne 2000 and in Amsterdam 2004, participants are expected to come from all areas of the social sciences, from economics, statistics, medicine, and natural sciences – and from all over the world. Keynote speakers in Naples will be Ludovic Lebart (Paris) and Willem Saris (Amsterdam), at the first and the last day of the conference we will have six half-day tutorials that are free of charge for conference participants. Here are the rules for paper givers; they are the same as for the previous RC33 meetings:

1. Every paper must contain a methodological problem (any area is fine, qualitative and/ or quantitative); applications are very welcome, however, they are not sufficient.
2. Speakers can only have one talk per session. This is also true for joint papers: it will not be possible for A and B to present one paper while B and A present another one during the same session. This would just extend the time allocated to a paper.
3. It is also not possible to present the same or a similar paper in two sessions. The organization committee will cross-check the sessions and inform the session organizers in doubtful cases.
4. We will have sessions of 80 minutes with four papers, and sessions of 120 minutes with six papers. It is up to the session organizer to allocate the time for discussion.
5. Please send your abstract either directly to the session organizer (see the list on the web site) or to Simona Balbi (Simona.Balbi@unina.it) who has the chair. Please do not send your abstract twice, for example, to two of the session organizers or to one of the session organizers and to Simona Balbi.

6. If you send your abstract to Simona Balbi and if the abstract sounds good, the organizing committee will offer the paper to one of the session organizers. The same holds for abstracts you send to the session organizers whose sessions are already full or where your paper does not fit in well. In these cases the session organizers will forward your abstract to the organization committee and we will forward it to a session we consider suitable for your topic.
7. Although the session organizers are responsible for their sessions, the final acceptance of a paper will take place after registration for the conference.
8. It will not be possible to present more than two papers; however, one paper is strictly preferred. If you are the co-author of a second or third paper, please take care that your co-author(s) register for the conference.
9. Please send your abstract as soon as possible to have a good chance that your preferred session has room available.
10. The abstract should not be longer than 250 words and it should indicate your name, your email address, your institutional affiliation and up to three keywords.

Deadline for abstracts: February 17, 2008

The session organizers or the organisation committee will inform you about the acceptance by the end of March at the latest.

Location: Università "Federico II" di Napoli, one of the most ancient universities in Europe (in 1224, the Holy Roman Emperor and King of Sicily Frederick II had founded the first publicly funded University in Naples), although the conference will take place in the modern campus known as Monte Sant'Angelo.

Date: September 1-5, 2008.

Monday afternoon and Friday morning will be devoted to tutorials.

Further information and updates:

<http://www.rc332008.unina.it/>

On behalf of the board of RC33 and the conference organization committee

SIMONA BALBI
ENRICA AMATURO
JÖRG BLASIUS
COR VAN DIJKUM

N. CARLO LAURO
ANNE RYEN
ROSANNA VERDE

Autorinnen und Autoren Jahrgang 1 (2007)

- Rolf Becker, Bern
- Henning Best, Mannheim
- Stefan Hunsicker, Mannheim
- Regula Imhof, Bern
- Ralf Künstler, Berlin
- Oliver Lipps, Neuchatel
- Britta Matthes, Nürnberg
- Guido Mehkop, Dresden
- Maike Reimer, München
- Kurt Salentin, Bielefeld
- Yvonne Schroth, Mannheim
- Wolfgang Sodeur, Duisburg-Essen

Rezensentinnen und Rezensenten Jahrgang 1 (2007)

- Nina Baur, Berlin
- Michael Braun, Mannheim
- Frank Faulbaum, Duisburg-Essen
- Franz Höllinger, Graz
- Jürgen H.P. Hoffmeyer-Zlotnik, Mannheim
- Natalja Menold, Mannheim
- Jost Reinecke, Bielefeld

Gutachterinnen und Gutachter Jahrgang 1 (2007)

Wir danken folgenden Kolleginnen und Kollegen, die uns für Jahrgang 1 (2007) der Methoden – Daten – Analysen Manuskripte begutachtet haben:

- Johannes Bacher, Linz
- Fred Berger, Zürich
- Jörg Blasius, Bonn
- Ingwer Borg, Mannheim
- Michael Bosnjak, Mannheim
- Josef Brüderl, Mannheim
- Edith de Leeuw, Utrecht
- Marc Deutschmann, Offenbach am Main
- Detlef Fetchenhauer, Köln
- Siegfried Gabler, Mannheim
- Anja Göritz, Erlangen
- Sabine Häder, Mannheim
- Ben Jann, Zürich
- Markus Klein, Köln
- Thomas Klein, Heidelberg
- Irene Kogan, Mannheim
- Ulrich Kohler, Berlin
- Caroline Kramer, Heidelberg
- Frauke Kreuter, Maryland
- Karin Kurz, Leipzig
- Oliver Lipps, Neuchatel
- Peter Lynn, Essex
- Natalja Menold, Mannheim
- Sören Petermann, Halle
- Bernhard Schimpl-Neimanns, Mannheim
- Sven Stadtmüller, Frankfurt
- Bernd Weiß, Köln

Hinweise für unsere Autorinnen und Autoren

Methoden – Daten – Analysen (MDA) veröffentlicht Beiträge aus dem Bereich der Empirischen Sozialforschung, insbesondere aus dem Bereich der Umfragemethodik. Im Vordergrund stehen Artikel, welche die methodischen und/oder statistischen Kenntnisse der Profession erweitern, sowie Beiträge, die sich mit der Anwendung der Methoden der Empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder solche, in denen ein statistisches Verfahren exemplarisch angewandt wird. Obwohl der Schwerpunkt auf Umfragemethoden liegen wird, sind Beiträge zu anderen methodischen Bereichen willkommen. Die Artikel sollen für eine breite Leserschaft von Wissenschaftlern und Praktikern im Bereich der Empirischen Sozialforschung verständlich sein.

Manuskripte, die bereits an anderer Stelle veröffentlicht sind oder gleichzeitig anderen Publikationsorganen zur Veröffentlichung angeboten worden sind, werden grundsätzlich nicht berücksichtigt. Eine spätere Veröffentlichung eines in der MDA erschienenen Beitrages ist möglich, sofern an exponierter Stelle auf die Ersterscheinung des Beitrages in der MDA hingewiesen wird.

Jeder Beitrag, der zur Veröffentlichung in MDA eingereicht wird, wird zunächst von den Herausgebern danach bewertet, ob er für eine Veröffentlichung grundsätzlich in Frage kommt.

Falls die Herausgeber einer Veröffentlichung grundsätzlich ablehnend gegenüber stehen, werden die Autoren unter Angabe von Gründen für diese Entscheidung informiert.

Falls die Herausgeber zur Ansicht gelangen, dass der Beitrag grundsätzlich zur Veröffentlichung in Frage kommt, wird er anonymisiert an mindestens zwei unabhängige Gutachter verschickt, die um eine Stellungnahme gebeten werden. Im Zweifelsfalle wird ein drittes Gutachten eingeholt.

Wird ein Beitrag nach Beschluss der Herausgeber in das Begutachtungsverfahren gegeben, erfolgt die abschließende Entscheidung über ein Manuskript auf der Basis der Gutachten durch die Herausgeber. Im Falle einer Ablehnung erhalten die Autoren eine ausführliche Begründung für die Ablehnung. Wird eine Überarbeitung eines Beitrages für erforderlich gehalten, erhalten die Autoren detaillierte Überarbeitungshinweise.

Unabhängig vom Ergebnis des Begutachtungsverfahrens werden die Autoren von der Entscheidung durch die Redaktion per E-Mail informiert.

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

Manuskripte müssen per E-Mail (mda@gesis.org) eingereicht werden. Der Umfang der Manuskripte soll inklusive Leerzeichen alles in allem nicht mehr als 70.000 Zeichen betragen.

Den Beiträgen sind Abstracts in Deutsch und Englisch (jeweils ca. 15 Zeilen) voranzustellen. Auch der Titel des Beitrages ist in Deutsch und Englisch einzureichen.

Um die Anonymität der Beiträge zu wahren, darf in einem Manuskript nur der Titel des Beitrages enthalten sein, nicht aber Namen oder Anschriften der Autoren; Name und Anschrift der Autoren müssen, gemeinsam mit dem Titel des Beitrages, auf einer separaten Seite eingereicht werden

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.

Tabellen enthalten Tabellenummer und Titel im Tabellenkopf, Abbildungen werden analog behandelt.

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.

Bei der Erstellung von Tabellen und Grafiken ist zu berücksichtigen, dass der Satzspiegel 11,5 cm (Breite) x 18,5 cm (Höhe) beträgt. Die Grafiken sind als jpeg- oder tif-Dateien zu liefern; Auflösung mindestens 150 dpi bei Schwarz-Weiss-Grafiken, 300 dpi bei farbigen Grafiken.

Die Beiträge sind unter Wahrung der gültigen Rechtschreiberegeln (neue Rechtschreibung) zu erstellen.

Werden in einem Beitrag empirische Daten verwandt, muss die Möglichkeit der Replikation bestehen. Im Falle einer Veröffentlichung in der MDA erklären sich die Autoren daher schriftlich bereit, Dritten auf deren Anfrage hin die Daten und Programmroutinen zur Verfügung zu stellen.

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.

Literaturhinweise im Text sind nach den folgenden Mustern aufzuführen: Müller (2002) – Müller (2002: 75) – (vgl. Müller 2002: 75) – (Müller 2002; Mayer/Müller/Schulze 2003).

Das Literaturverzeichnis ist wie folgt zu gestalten:

Buchveröffentlichungen:

Strobl, R. und W. Kühnel, 2000: Dazugehörig und ausgegrenzt. Analysen zu Integrationschancen junger Aussiedler. Weinheim/München: Juventa.

Zeitschriftenbeiträge:

Fuchs, M. und M. Sixt, 2007: Zur Nachhaltigkeit von Bildungsaufstiegen. Soziale Vererbung von Bildungserfolgen über mehrere Generationen. Kölner Zeitschrift für Soziologie und Sozialpsychologie 59: 1-29.

Beiträge in Büchern:

Braun, M. und I. Borg, 2004: Berufswerte im zeitlichen und im Ost-West-Vergleich. S. 179-199 in: R. Schmitt-Beck, M. Wasmer und A. Koch (Hg.): Sozialer und politischer Wandel in Deutschland. Analysen mit ALLBUS-Daten aus zwei Jahrzehnten. Wiesbaden: VS-Verlag für Sozialwissenschaften.

Internetquellen:

Stadtmüller, S. und R. Porst, 2005: Zum Einsatz von Incentives bei postalischen Befragungen. ZUMA How-to-Reihe, Nr. 14 (Mannheim: Zentrum für Umfragen, Methoden und Analysen, ZUMA). http://www.gesis.org/Publikationen/Berichte/ZUMA_How_to/Dokumente/pdf/how-to14rp.pdf (29.06.2007).