

werden. Ein Access Panel kann daher nur Personen ent-
lich befragungsbereit sind

Uwe Engel
Simone Bartsch
Christiane Schnabel
Helen Vehre

Wissen- schaftliche Umfragen

Methoden und Fehlerquellen

campus

darstellt.
efonische Festnetz um eine Rekrutierung über Mobilfunk ergänzt.
Man kann sich dies über eine Analogie bei der Ziehung

Uwe Engel, Dr. phil., ist Professor und leitet das Sozialwissenschaftliche Methodenzentrum der Universität Bremen.

Simone Bartsch, Dr. rer. pol., ist Survey-Managerin in der Abteilung Sozio-oekonomisches Panel am DIW Berlin.

Diplom-Soziologin *Christiane Schnabel* leitet die wissenschaftliche Abteilung der PARSHIP GmbH, Hamburg.

Helen Vehre, Dr. rer. pol., ist Senior Projektleiterin am LINK Institut für Markt- und Sozialforschung, Frankfurt.

Uwe Engel, Simone Bartsch, Christiane Schnabel,
Helen Vehre

Wissenschaftliche Umfragen

Methoden und Fehlerquellen

Campus Verlag
Frankfurt/New York

© Campus Verlag GmbH

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie.

Detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-593-39757-3

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung des Verlags unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Copyright © 2012 Campus Verlag GmbH, Frankfurt am Main

Umschlaggestaltung: Guido Klütsch, Köln

Satz: Publikations Atelier, Dreieich

Druck und Bindung: CPI buchbücher.de, Birkach

Gedruckt auf Papier aus zertifizierten Rohstoffen (FSC/PEFC).

Printed in Germany

Dieses Buch ist auch als E-Book erschienen.

www.campus.de

Inhalt

Vorwort	9
1. Einführung	12
2. Umfrageteilnahme: Theorie und Forschungsstand	20
2.1 Allgemeine Theorien zur Umfrageteilnahme	20
2.2 Besonderheiten der Umfrageteilnahme nach Modus	27
2.3 Antwortneigung	30
2.4 Einfluss von Interviewern	33
3. Umfragedesign	38
3.1 Rekrutierung	39
3.2 Konzeption und Frageprogramm der Rekrutierungsinterviews	49
3.3 Panelaufbau	54
4. Stichprobe: Plan und Realisierung	59
4.1 Auswahlrahmen	59
4.2 Antwortraten: Rekrutierung, Wiederbefragungsbereitschaft, Panelteilnahme	66
4.3 Umfrageklima	68
4.4 Wiederbefragungsbereitschaft und Panelteilnahme	74
4.5 Weitere Rekrutierungs- und Panelaufbauphasen	77
4.6 Panelsurveys	81
4.7 Verteilung soziodemographischer Variablen im Vergleich von Rekrutierungsstichproben, Access Panel und Mikrozensus 2009	87

6	Wissenschaftliche Umfragen	
5.	Der Kontaktverlauf als Determinante der Umfrageteilnahme	97
5.1	Einführung	97
5.2	Zielgrößen und Verlaufstypologie	101
5.3	Interviewwahrscheinlichkeit und Kontaktverlauf	112
5.4	Incentives	128
5.5	Persönlicher Bezug und Sensitivität der Umfragethematik	129
5.6	Flexibles Umfragedesign	131
5.7	Verweigerungswahrscheinlichkeit	135
5.8	Wahrscheinlichkeit, einen Haushalt nicht (wieder) zu erreichen	139
6.	Überzeugung zögerlicher Personen: Aufwand und Strategien	142
6.1	Generelle Befragungsbereitschaft	142
6.2	Überzeugungsstrategien	150
7.	Umfrageerfahrung und Teilnahmegründe	156
7.1	Umfrageerfahrung	156
7.2	Teilnahmegründe	160
8.	Integration und Soziodemographie	165
9.	Interviewereffekte	170
9.1	Interviewwahrscheinlichkeit als zu erklärende Zielgröße	170
9.2	Stimmliche und kommunikative Merkmale	173
9.3	Antwortneigung und stimmliche und kommunikative Merkmale	178
9.4	Interviewwahrscheinlichkeit und Einstellungen von Interviewern	182
9.5	Interview- und Verkaufserfahrung der Interviewer	195
9.6	Geschlecht und Alter des Interviewers	197
10.	Modellierung der Antwortneigung	201

11. Panelteilnahme	215
11.1 Das Modell in der Übersicht	215
11.2 Zielgröße 1: Teilnahme am Rekrutierungsinterview	216
11.3 Zielgröße 2: Wiederbefragungsbereitschaft	217
11.4 Zielgröße 3: Panelteilnahme	230
12. Soziale Integration und Panelteilnahme	254
12.1 Integrationsindikatoren und Wiederbefragungsbereitschaft	254
12.2 Verhältnis zur Politik	256
12.3 Ehrenamtliches Engagement	261
12.4 Interesse für wissenschaftliche Themen	262
12.5 Wertrationale Einstellungen	263
12.6 Leben in Deutschland	272
12.7 Normakzeptanz	273
12.8 Integration und sozialer Austausch	273
12.9 Soziodemographie	279
13. Befragungseffekte	286
13.1 Die Experimente und Modus-Vergleiche in der Übersicht ..	286
13.2 Auswertungstechniken	289
13.3 Skalen- und Fragerichtung	291
13.4 Skalenart	305
13.5 Fünf- vs. siebenstufige Zustimmungsskalen	306
13.6 Elfstufige Zustimmungsskalen	311
13.7 Skalenlabel	321
13.8 Fragereihenfolgeeffekte	335
13.9 Siebenstufige Skalen im Modus-Vergleich	344
13.10 Zusammenfassung	350
14. Im Rückblick	355
Literatur	360

Vorwort

Zum 1. Januar 2008 startete das von meinem Arbeitsgebiet »Statistik und empirische Sozialforschung« der Universität Bremen koordinierte Schwerpunktprogramm 1292 »Survey Methodology« der Deutschen Forschungsgemeinschaft (www.survey-methodology.de).

In vorliegender Monographie berichten wir über Ergebnisse aus einem dieser Schwerpunktprojekte, dem »Access Panel and Mixed Mode Internet Survey«. Es startete zeitgleich mit dem Schwerpunktprogramm und realisierte mit dem zufallsrekrutierten Aufbau des Access Panels den größten Teil seiner Arbeit in den ersten beiden Programmjahren. Seitdem realisiert es Befragungen und Methodenexperimente innerhalb dieses Panels.

Die Studie wurde vom Projektstandort »Bremen« aus realisiert. Dort fand auch ein kleinerer Teil der Befragungen statt. Der größere Teil der Befragungen wurde für die Projektgruppe vom Sozialwissenschaftlichen Umfragezentrum (SUZ) am Standort »Duisburg« realisiert.

Der Projektgruppe gehörten neben mir meine seinerzeitigen wissenschaftlichen Mitarbeiterinnen Simone Bartsch, Christiane Schnabel und Helen Vehre an, außerdem die Leiterin meines Sekretariats, Sabine Sommer.

Ich bin meinem Team für die tatkräftige Mitwirkung beim Aufbau des Access Panels sehr dankbar. Gleiches gilt für die studentischen Mitarbeiterinnen und Mitarbeiter des Projekts. Außerordentlich hilfreich war die Unterstützung, welche das Projekt durch die Universität Bremen erfahren hat.

Bei der Deutschen Forschungsgemeinschaft bedanke ich mich für die finanzielle Förderung des Vorhabens.

Wesentliche Analysen, über die in vorliegender Monographie berichtet wird, wurden zuvor auf wissenschaftlichen Tagungen zur Diskussion gestellt. Den Anfang machte das in Kapitel 10 vorgestellte Modell zur Schätzung der anfänglichen Interviewwahrscheinlichkeit. Es wurde im Vorfeld der Haupterhebung gestützt auf Ergebnisse aus der Pilotstudie, die Anfang desselben Jahres in Hamburg realisiert worden war, auf der Herbsttagung 2008 des Schwerpunktprogramms am 25./26. September an der Universität Kassel vorgestellt.

Gegenstand des Vortrags auf der Frühjahrstagung 2009 des Schwerpunktprogramms am 6./7. Mai in Bremen waren die Ergebnisse aus unseren Pilotstudien

und die Konsequenzen, die daraus für Design und Frageprogramm der zu jener Zeit gerade im Feld befindlichen Haupterhebung sowie für die Analyse der Wiederbefragungsbereitschaft zu ziehen waren.

Auf der internationalen Tagung des Schwerpunktprogramms »Improving Survey Methods«, die am 12. und 13. November 2009 in Bremen veranstaltet wurde, wurde das Modell zur Erklärung der anfänglichen Interviewwahrscheinlichkeit, hier nochmals gestützt auf Auswertungen der Hamburg-Pilotstudie, einem größeren Fachpublikum vorgestellt. Außerdem wurden erste Auswertungen von am Projektstandort »Bremen« erhobenen Daten der Haupterhebung zur eingesetzten *Survey Attitude Scale* sowie zur Erklärung der Wiederbefragungsbereitschaft auf der Basis dieser Skala und der erhobenen Para- und Metadaten zur Diskussion gestellt.

Über Ergebnisse der Haupterhebung wurde sodann umfassend in der Statistischen Woche der Deutschen Statistischen Gesellschaft am 17. September 2010 sowie auf der Herbsttagung 2010 des Schwerpunktprogramms am 6. Oktober in Bremen berichtet. Ein Schwerpunkt lag dabei nochmals auf dem in Kapitel 10 vorgestellten Modell, nun gestützt auf Auswertungen der betreffenden Daten der Haupterhebung. Da sich das Modell auf die unten in den Kapiteln 5 bis 8 vorgestellten Faktoren zur Erklärung der anfänglichen Interviewwahrscheinlichkeit stützt (Kontaktverläufe, Überzeugungsstrategien, Umfrageerfahrung, Teilnahme Gründe), waren entsprechende Analysen Bestandteil der beiden Vorträge. Außerdem wurden erste Auswertungen zur Erklärung der Wiederbefragungsbereitschaft und der Panelteilnahme vorgestellt, im Vortrag für die Herbsttagung des Schwerpunkts ergänzt um erste Auswertungen der Befragung der Interviewer/innen.

Unser Ansatz zur Verwendung von Paradata in Mehrebenenmodellen zur Erklärung der Umfrageteilnahme war auch Gegenstand einer Posterpräsentation auf dem Paradata Symposium in London am 9. Dezember 2010.

Im Frühjahr 2011 stand die Frage der Determinanten der Panelteilnahme auf der Agenda eines Vortrags auf der General Online Research Tagung am 15. März an der Universität Düsseldorf sowie auf der Tagung der European Survey Research Association in Lausanne am 20. Juli 2011.

Auf der Basis des Feedbacks auf die diversen Tagungspräsentationen wurde das Gesamtmodell zur Erklärung der anfänglichen Teilnahme am Rekrutierungsinterview, der Wiederbefragungsbereitschaft und der faktischen Panelteilnahme sukzessive optimiert und zu den Fassungen weiterentwickelt, die wir in vorliegender Monographie präsentieren. Dies sind zugleich die Fassungen, die auf drei zurückliegenden Tagungen des Jahres 2011 präsentiert wurden. Zum einen handelt es sich um zwei Vorträge, die am 12./13. Mai auf der Frühjahrstagung 2011 des Schwerpunktprogramms in Bremen zu den Themenfeldern Panelteilnahme und Interviewereffekte gehalten wurden. Zum zweiten handelt es sich um zwei Vorträge, die zu denselben Themenfeldern auf der Spring Methodology Confe-

rence der American Sociological Association am 19./20. Mai 2011 in Tilburg gehalten wurden. Drittens wurde in einem Vortrag für die Statistische Woche der Deutschen Statistischen Gesellschaft, der am 21. September 2011 an der Universität Leipzig gehalten wurde, ein Modell zur Messung von Wohlbefinden präsentiert, deren 11-stufige Skalen für die Komponenten Glück, Zufriedenheit und Zuversicht in Kapitel 13 auf mögliche Befragungseffekte untersucht werden. Das dort skizzierte Modell zur Messung von Lebensqualität fand in einer leicht veränderten Variante auch Eingang in die sich über drei Wellen erstreckende Panelanalyse, die wir auf der Frühjahrstagung der Sektionen Methoden der empirischen Sozialforschung und Modellbildung und Simulation der Deutschen Gesellschaft für Soziologie am 30. März 2012 an der Universität Bremen präsentiert haben. Vorausgegangen war dieser Panelanalyse eine Analyse, in der das Teilnahmeverhalten über die drei Wellen hinweg daraufhin untersucht wurde, ob es davon berührt wird, ob Teilnehmer/innen in der vorgelagerten Rekrutierungsphase erst von einer Teilnahme überzeugt werden wollten. Diese Panelanalyse war auf der 2. Internationalen Tagung des PPSM Schwerpunktprogramms vorgestellt worden, die am 17./18. November 2011 in Bremen stattgefunden hatte.

Prof. Dr. Uwe Engel

Im Juni 2012

1. Einführung

Ein Access Panel stellt einen Pool von Adressen von Personen dar, die sich dazu bereit erklärt haben, zu Umfragen dieses Access Panels eingeladen zu werden. Ein Access Panel kann daher nur Personen enthalten, die grundsätzlich befragungsbereit sind.

Die Adressen dieser befragungsbereiten Personen können Telefonnummern oder E-Mail-Adressen sein. Sie repräsentieren den Auswahlrahmen, aus dem heraus die für Befragungen erforderlichen Stichproben gezogen werden. Von Bedeutung ist daher die Frage, wie geeignet der Auswahlrahmen für diesen Zweck ist.

Wie diese Frage zu beantworten ist, hängt im Prinzip von drei Faktoren ab. Zum ersten davon, wie der Auswahlrahmen entstanden ist. Zum zweiten davon, für welche Grundgesamtheit der Auswahlrahmen Verwendung finden soll. Und zum dritten davon, wie Ausfälle das Rekrutierungsergebnis verändern.

Der erste Faktor kann auch so formuliert werden: Wie sind die Befragungspersonen in das Access Panel und damit den Auswahlrahmen gelangt? Entscheidend ist dabei, ob die Rekrutierung in das Access Panel auf der Basis eines kontrollierten Auswahlverfahrens erfolgt ist oder nicht.

Bei kleinen Populationen wird nicht selten der Fall eintreten, dass alle Populationselemente ausgewählt werden, um den Auswahlrahmen herzustellen. In aller Regel sind dies zugleich spezielle Populationen. Studierende eines Studiengangs einer Universität wären hier als ein Beispiel anzuführen, bei denen die betreffende Studierendendatei quasi als Access Panel genutzt wird, um daraus gegebenenfalls wiederholt Stichproben für Befragungen zu ziehen. Analog könnte beispielsweise bei Belegschaften von Betrieben oder den Mitgliedern eines Verbandes verfahren werden.

Bei großen Populationen kann der Regelfall nur sein, dass nicht alle Populationselemente den Auswahlrahmen bilden. Man denke nur an die wahlberechtigte Bevölkerung Deutschlands als Beispiel einer großen allgemeinen Population. Sehr groß können aber auch spezielle Populationen sein, in denen Umfragen realisiert werden sollen. Beispiele wären Befragungen unter Angehörigen bestimmter Berufsgruppen, Branchen oder ethnischer Gruppen. Dann ist unvermeidlich zu fragen, ob der Erstellung des späteren Auswahlrahmens für das Access Panel selbst ein kontrollierter Auswahlprozess von Populationselementen zugrunde liegt und

wie sich das Verhältnis des Auswahlrahmens zur Population in punkto Über- bzw. Unterdeckung darstellt.

Man kann sich dies über eine Analogie bei der Ziehung von Zufallsstichproben verdeutlichen. Solche Stichproben werden in aller Regel direkt aus einer Auswahlgesamtheit gezogen, welche die fragliche Population repräsentieren soll. Es ist dann stets zu beachten, dass alle Populationselemente eine von Null verschiedene, angebbare Ziehungswahrscheinlichkeit besitzen. In gleicher Weise gilt dies bei Access Panels für den Rahmen, der nachgelagert als Grundlage für die Ziehung von Stichproben innerhalb des Access Panels dient. Der Auswahlrahmen tritt gewissermaßen zwischen die Population auf der einen Seite und die Stichprobe auf der anderen Seite. Die Ziehung von Zufallsstichproben innerhalb eines Access Panels kann dabei die speziell von Zufallsstichproben erwarteten Vorteile dann nicht mehr erfüllen, wenn der Zufallscharakter der Auswahl zuvor bereits bei der Erstellung des Auswahlrahmens zerstört oder beschädigt wurde.

Grundsätzlich kann dies auf unterschiedliche Weise geschehen. Zu aller erst dadurch, dass überhaupt auf eine kontrollierte Auswahl verzichtet wird. Dies ist stets bei passiv rekrutierten Access Panels der Fall, da es dort einzig und allein davon abhängt, ob jemand auf die Befragungsgelegenheit eines Access Panels aufmerksam wird und sich bei Interesse für eine Mitwirkung daran entscheidet. Aber auch bei aktiv rekrutierten Access Panels kann es zu Beeinträchtigungen des Zufallscharakters der Auswahl kommen, und zwar aus in der Sozialforschung hinlänglich bekannten Gründen. Zum einen kann es zu systematischen Ausfällen bei dem Versuch kommen, einen Stichprobenplan zu realisieren. Es resultiert trotz Zufallsziehung dann eine verzerrte Stichprobe. Wird diese Stichprobe nun aber als Rekrutierungsstichprobe für die Erstellung des Auswahlrahmens des Access Panels genutzt, so überträgt sich diese Verzerrung auf diesen Rahmen und die später daraus gezogenen Stichproben. Wer verstehen will, in welcher Weise der Auswahlrahmen eines Access Panels gegebenenfalls verzerrt ist, muss auch verstehen, in welcher Weise die vorgelagerten Rekrutierungsstichproben verzerrt sind. Und dies bedeutet zu verstehen, von welchen Faktoren es abhängt, wer an Umfragen teilnimmt und wer nicht. Genau diese Frage steht im Mittelpunkt vorliegender Ausarbeitung.

Zum anderen kann der Zufallscharakter einer Rekrutierungsstichprobe dadurch beeinträchtigt werden, dass die verwendete Auswahlgesamtheit die Grundgesamtheit nur teilweise abdeckt. In Deutschland kann dieser Fall inzwischen bei telefonischen Befragungen eintreten, wenn sich die Auswahl auf Festnetzanschlüsse beschränkt und dadurch diejenigen ausgeklammert werden, die nur oder in erster Linie nur noch über Mobilfunk erreicht werden können. Deshalb wurde in vorliegendem Projekt die Rekrutierung über das telefonische Festnetz um eine Rekrutierung über Mobilfunk ergänzt.

Access Panels stellen durchaus keine Seltenheit dar. Bei einer Internetrecher-

che im Sommer 2010 konnten international 139 Access Panels gefunden werden, davon 54 in Deutschland und 60 in den USA. Wir haben über die Ergebnisse dieser Recherche an anderer Stelle genauer berichtet (Vehre 2011). Bemerkenswert ist, dass gut 60 Prozent dieser recherchierten Access Panels passiv rekrutiert sind. Das bedeutet, dass der Anmeldung zum Panel kein wissenschaftlich kontrolliertes Auswahlverfahren zugrunde liegt. Stattdessen werden Personen beispielsweise durch Werbefbanner auf Webseiten auf die Möglichkeit der Teilnahme an einem Access Panel aufmerksam und entscheiden sich bei Interesse gegebenenfalls für eine Mitwirkung daran. Zudem gibt es Webseiten, deren Zweck darin besteht, interessierten Personen einen Überblick über Teilnahmemöglichkeiten an Access Panels zu vermitteln. Dass solche Seiten funktionieren, hängt wesentlich damit zusammen, dass der Überblick über Teilnahmemöglichkeiten zugleich einen Überblick über Verdienstmöglichkeiten darstellt. Umfrageteilnahme gegen Bezahlung ist das tragende Prinzip.

Knapp 10 Prozent der recherchierten Access Panels erwiesen sich hingegen als aktiv rekrutiert und weitere circa 15 Prozent als teilweise aktiv und teilweise passiv. Bei aktiver Rekrutierung werden die Zielpersonen direkt zur Teilnahme aufgefordert, beispielsweise telefonisch oder per E-Mail. Damit verbindet sich der Vorteil, der Aufforderung zur Teilnahme an einem Access Panel ein wissenschaftlich kontrolliertes Auswahlverfahren zugrunde zu legen. Dies ist der Weg, der in vorliegender Studie mit dem Einsatz telefonischer Zufallsauswahlen eingeschlagen wurde.

Rund 78 Prozent der recherchierten Access Panels erwiesen sich als reine Onlinepanels, bei weiteren 11 Prozent ergänzt um einen weiteren Erhebungsmodus. Geschätzt wird, dass sich die Zahl der Onlinepanels weltweit auf über 600 beläuft (Batinic/Moser 2005).

Ausfälle und Befragungseffekte als Fehlerquellen

Ein großer Vorteil von Zufallsstichproben besteht darin, auf ihrer Grundlage berechnete Kennwerte zur Schätzung der betreffenden Werte in der Grundgesamtheit (Population) verwenden zu können, aus der die Stichprobe gezogen wurde. So ist man in aller Regel daran interessiert, die Grundgesamtheit in der einen oder anderen Hinsicht zu charakterisieren, hat dafür aber nur die Informationen aus der Zufallsstichprobe verfügbar. So mag von Interesse sein, den Anteil von Wählern einer bestimmten politischen Partei zu kennen. Sofern im Zuge der Stichprobenrealisierung keine systematischen Ausfälle auftreten, ist dies eine leicht zu lösende Standardaufgabe, da der Anteilswert in der Grundgesamtheit über den Anteilswert in der Stichprobe geschätzt werden kann. Für den Fall systematischer Ausfälle gilt dies allerdings nicht so ohne weiteres. Denn wenn sich die Antworten, welche die an der Umfrage teilnehmenden Personen geben, von

den Antworten unterscheiden, welche die nicht-teilnehmenden Personen auf dieselbe Frage gegeben hätten, wenn sie denn an der Umfrage teilgenommen hätten, dann kann der aus den Angaben der Teilnehmer/innen berechnete Anteilswert den gesuchten Anteilswert in der Grundgesamtheit nur verfehlen. Er wird je nach Richtung des Unterschieds in den Antworten von Teilnehmern und Nichtteilnehmern größer oder kleiner als der wahre Wert ausfallen.

Nun ist dieser Unterschied im Allgemeinen unbekannt, liegt es doch in der Natur der Sache, dass in einer Umfrage Antworten nur von teilnehmenden Personen vorliegen können. Das Problem besteht ja gerade darin, dass von nicht-teilnehmenden Personen die Antworten fehlen.

Das macht es schwierig, die Verzerrung einer Stichprobe in Bezug auf diejenigen Antwortverteilungen einzuschätzen, die in einer Umfrage auf die dort gestellten inhaltlichen Fragen abgegeben werden. Zwar besteht im Allgemeinen die Möglichkeit, Verteilungen einer Stichprobe mit externen Referenzverteilungen abzugleichen und diesen Abgleich zur Grundlage einer korrigierenden Stichprobengewichtung zu nehmen. Typische Beispiele dafür dürften Vergleiche mit dem Mikrozensus oder den Registerdaten von Einwohnermeldeämtern sein. Jedoch kann diese Option nicht darüber hinwegtäuschen, dass das für diesen Vergleich verfügbare Set von Variablen stark begrenzt ist. Das wäre zwar solange keine echte Restriktion, wie das verfügbare Set von externen Register- bzw. Referenzvariablen hoch mit den Variablen korrelieren, die in einer Umfrage inhaltlich erhoben werden. Aber genau das kann im Allgemeinen nicht unterstellt werden und hat für diesen Fall zur Folge, dass eine kompensatorische Stichprobengewichtung nur partiell erfolgreich sein kann. Sie kann zwar den Verzerrungsgrad einer Stichprobe in Bezug auf die externen Referenzverteilungen minimieren, ohne dass dies jedoch implizieren würde, damit auch etwaige Verzerrungen in Bezug auf die Antwortverteilungen zu den inhaltlichen Variablen der Umfrage zu verringern.

In vorliegendem Projekt wurde daher versucht, Informationen auch von Personen zu erheben, die unter normalen Umständen nicht an einer Umfrage teilnehmen und folglich Ausfälle darstellen würden. Dazu wurden den Personen, die um ein Interview gebeten wurden, neben der Vollfassung des Interviews gegebenenfalls auch eine kürzere Kern- und eine noch kürzere Exitfassung angeboten. Zudem wurde versucht, zögerliche Personen davon zu überzeugen, das erbetene Interview zu geben.

Damit stehen Informationen zur Verfügung, die aus dem Umfrageprozess selbst stammen und es erlauben, Vergleiche in inhaltlichen Variablen zwischen unterschiedlich befragungsbereiten Personen anzustellen. Ergänzt werden diese Informationen durch Informationen über das Interview und den Interviewteilnehmer. Zum einen handelt es sich dabei um Einschätzungen, die vom Interviewer auf einer Reihe von zuvor entwickelten Skalen abzugeben waren (Paradaten) und zum anderen um Einschätzungen der befragten Personen selbst (Metadaten).

Eine relevante Teilmenge dieser vom Interviewer vorzunehmenden Einschätzungen liegt dabei auch für Personen vor, die trotz aller Bemühungen nicht zur Teilnahme am Interview zu bewegen waren.

Die Interviewer bzw. Interviewerinnen werden in diesem Umfragedesign selbst zu einer relevanten Größe. Wir haben sie daher in einer der Haupterhebung unmittelbar vorgelagerten Befragung selbst um eine Reihe von Auskünften gebeten, um später für die Haupterhebung etwaige Interviewereffekte identifizieren zu können (siehe Kapitel 9).

Nun können Umfrageergebnisse nicht nur durch systematische Ausfälle (Nonresponse-Effekte) verfälscht werden. Ein weiterer Komplex von Einflussfaktoren ist in der Art und Weise zu sehen, in denen die Fragen einer Umfrage gestellt werden. Ideal wäre es, wenn vom *modus operandi* keinerlei Einfluss auf Antworten ausgehen würde. Denn im Prinzip sollen die Antworten, die in einer Umfrage auf dort gestellte Fragen abgegeben werden, die betreffende Wirklichkeit widerspiegeln und nicht erst durch Modalitäten der Befragung selbst erzeugt worden sein. Zumindest ist anzustreben, den speziell auf Frage- und Antwortformate zurückführbaren Eigenanteil in der Variation von Antworten auf Fragen einer Umfrage identifizieren und von übrigen Variationsursachen trennen zu können.

Wie aber können Antworten durch Fragen erst erzeugt, wie bestimmte Antworten begünstigt oder benachteiligt werden? Aus dem Spektrum möglicher Einflussfaktoren wollen wir uns in vorlegender Ausarbeitung auf den Faktor »Fragerichtung« sowie unterschiedliche Möglichkeiten konzentrieren, Antwortskalen vorzugeben. Außerdem sind mögliche Fragereihenfolgeeffekte zu bedenken. In die 20-minütigen Rekrutierungsstichproben für das Access Panel sowie in nachgelagerte Umfragen innerhalb des Panels wurde daher eine Reihe von Experimenten eingebaut, um Effekte identifizieren zu können, die auf Unterschiede in Frage- und Antwortformaten zurückzuführen sind.

Determinanten der Umfrageteilnahme

Die Wahrscheinlichkeit einer Umfrageteilnahme hängt von einer ganzen Reihe von Faktoren ab, die sich nach unterschiedlichen Gesichtspunkten klassifizieren lassen.

Eine erste Klassifikationsmöglichkeit liegt in der Unterscheidung, ob Faktoren primär die Kontakt- oder Kooperationswahrscheinlichkeit beeinflussen. Um ein Interview zu erhalten, ist es erforderlich, Kontakt zu den Zielpersonen herstellen zu können. Die betreffenden Personen müssen erreicht werden können. Bei telefonischen Befragungen über das Festnetz bedeutet dies zunächst den Kontakt mit dem Haushalt herstellen zu können. Gelingt dies, wäre als nächstes die Zielperson auszuwählen. Dies kann, muss aber nicht die Person sein, die das Ge-

spräch entgegennimmt. Folglich mag es bereits auf dieser Stufe von der Kooperationsbereitschaft einer Kontaktperson abhängen, ob die Zielperson ausgewählt werden kann und ob der Kontakt zu ihr vermittelt wird. Gelingt auch dies, hängt es entscheidend von der Kooperationsbereitschaft der Zielperson selbst ab, ob sie das erbetene Interview gibt. Die Frage der Erreichbarkeit bezieht sich also sowohl auf den Haushalt als nachgelagert auch auf die auszuwählende Person. Dabei kann es zu unterschiedlichen Kontaktverläufen kommen. Kapitel 5 zeigt auf, wie diese die Wahrscheinlichkeit beeinflussen, dass ein Interview zustande kommt.

Kommt eine Gesprächsverbindung zustande, so bedeutet dies per se noch nicht, dass auch ein Gespräch zustande kommt. Ein nicht unerheblicher Teil von kontaktierten Haushalten verweigert eine Mitwirkung am Gespräch so spontan, dass der Interviewer bzw. die Interviewerin nicht einmal den Einleitungssatz zu Ende vorlesen kann. Für diese Fälle nutzt die beste Interviewerschulung nichts. Leicht abgestuft gibt es sodann eine weitere Gruppe von Haushalten, bei denen sich die Person, die das Gespräch entgegennimmt, zwar die Einleitung anhört, dann aber ohne zu zögern und ohne mit sich reden zu lassen, die Bitte nach einem Interview ablehnt. Auch hier wird dem Interviewer praktisch keine Gelegenheit gegeben, positiv auf die Kooperationsbereitschaft der Person einzuwirken. Auf der anderen Seite gibt es natürlich auch den Fall, dass sich der kontaktierte Haushalt befragungsbereit zeigt. Zwischen diesen beiden Extremen spannt sich allerdings ein Kontinuum zögerlicher Personen auf, die eine Interviewteilnahme für sich weder von vornherein ausschließen noch von vornherein akzeptieren, und in der Kontaktsituation folglich mit sich reden lassen. Auf solcherart zögerliche Personen können dann auch Bemühungen von Seiten der Interviewer/innen gerichtet werden, etwaige Bedenken zu zerstreuen, um auch zögerliche Personen von einer Teilnahme am Interview zu überzeugen. Im Kern geht es dabei darum, über Überredungs- bzw. Überzeugungsversuche die Kooperationswahrscheinlichkeit positiv zu beeinflussen. Kapitel 6 informiert über unsere diesbezüglichen Ergebnisse.

Sowohl die Kontakt- als auch die Kooperationswahrscheinlichkeit kann auf Faktoren beruhen, die sich danach klassifizieren lassen, ob sie primär in den zu befragenden Personen und deren Lebensumständen, den Umständen der Kontakt- und Interviewsituation selbst oder Elementen des Umfragedesigns begründet sind. Dies ist eine zweite Klassifikationsmöglichkeit. So ist die Erreichbarkeit eines Haushaltes bzw. einer Zielperson zum Beispiel eine Frage der Zeiten berufsbedingter Abwesenheit, ebenso aber auch eine Frage der Tageszeiten, zu denen die Kontaktversuche stattfinden. Letzteres ist ein Element des Umfragedesigns und liegt somit unter der Kontrolle des Sozialforschers, erstgenanntes nicht. In vergleichbarer Weise kann die Kooperationsbereitschaft von früheren Erfahrungen der angefragten Personen mit Sozial- oder Marktforschungsstudien oder beispielsweise ihren Einstellungen gegenüber der Umfrageforschung abhängen und

sich insoweit aus Umständen ergeben, die primär in der jeweiligen biografischen Erfahrung der angefragten Personen liegen. Die Kooperationsbereitschaft kann aber ebenso auch von Faktoren wie der Umfragethematik, der Länge des Interviews oder des Umfragemodus und somit von Elementen abhängen, die dem Umfragedesign zuzurechnen sind. Eine Rolle kann auch die Institution spielen, für die die Umfrage realisiert wird. Als von nicht minder wichtiger Bedeutung kann sich aber auch die Kontakt- und Interviewsituation selbst herausstellen. Eine Teilnahme kann auf Gründen beruhen, die sich aus der Kontaktsituation selbst ergeben, etwa weil sich Personen einfach nett gebeten fühlen. Auch mag eine Interviewanfrage erfolgreich sein, weil die angefragte Person gerade Zeit hat und eine Teilnahme somit keine Opportunitätskosten erzeugt. Teilnahmegründe können aber auch weniger situativ als vielmehr stärker in der Vorgeschichte von angefragten Personen verankert sein. Wir untersuchen die Faktoren Umfrageerfahrung und Teilnahmegründe genauer in Kapitel 7.

Ist dann erst einmal ein Interview zustande gekommen, sind weitere situative Faktoren in Betracht zu ziehen, etwa die seitens der befragten Personen empfundene Interviewatmosphäre. Gleiches gilt für den Grad empfundener Sensitivität der Fragen und anderer Aspekte des Interviews. Determinanten dieser Art können im vorliegenden Kontext zwar nicht erklären, warum jemand am Rekrutierungsinterview selbst teilnimmt, wohl aber, ob er oder sie bereit ist, sich erneut befragen zu lassen. Schwerpunktmäßig gehen wir in Kapitel 11 auf die Fragen von Wiederbefragungsbereitschaft und Panelteilnahme ein.

Zu den Faktoren, die primär in den zu befragenden Personen und deren Lebensumständen verankert sind, zählen zum einen die schon angesprochenen Einstellungen gegenüber der Umfrageforschung. Zwar gelten Einstellungen im Allgemeinen nicht als gute Verhaltensprädiktoren; diese Aussage relativiert sich allerdings dann, wenn es nicht um allgemeine, sondern um spezielle Einstellungen gegenüber konkreten Einstellungsobjekten geht. Entsprechend stark sind unsere betreffenden Analysen auf Einstellungen gegenüber Umfrageforschung fokussiert. Ihr Einfluss wird vor allem in Kapitel 11 beleuchtet.

Über ein weiteres Set von Faktoren berichten wir in den Kapiteln 8 und 12. Einerseits geht es dabei um Merkmale der Soziodemographie und somit um Effekte, die auf die Platzierung einer Person in der Sozialstruktur zurückzuführen sind. Zum anderen geht es um verschiedene Faktoren, die soziologisch betrachtet unter den Begriff der sozialen Integration subsumiert werden können. Wir greifen damit eine erkenntnisleitende Idee aus der einschlägigen Literatur auf. Danach steigt die Wahrscheinlichkeit einer Umfrageteilnahme mit der Integration einer Person in die umgebende Gesellschaft. Dies kann auf unterschiedliche Weise geschehen. Beispielsweise über Vereinsmitgliedschaften und die Übernahme ehrenamtlicher Tätigkeiten. Oder durch die Ausprägung von Bindungen. So kann sich jemand dem Land, in dem er oder sie lebt, mehr oder weniger stark

verbunden fühlen. Ebenso denkbar und in Betracht zu ziehen sind gesellschaftsaffine Einstellungen. Mit der Gesellschaft, in der man lebt, kann man sich auch dadurch identifizieren, dass man ihre basalen Regeln und vorherrschende normativen Vorstellungen übernimmt. Oder durch politische Partizipation, der Beteiligung an Wahlen beispielsweise. In vorliegender Studie werden daher eine Reihe solcher Integrationsindikatoren in die Analyse einbezogen.

Vorliegende Ausarbeitung gliedert sich in drei größere Teile. Kapitel 1 und 2 führen in die Thematik ein und liefern einen knappen Überblick über Theorien und Forschungsstand zur Frage der Umfrageteilnahme. Die Kapitel 3 und 4 informieren über das Design der Studie und seine Realisierung. Es schließen sich die Kapitel an, in denen über die vorgenommenen Auswertungen berichtet wird. Kapitel 5 bis 10 ist gemeinsam, dass sie die Wahrscheinlichkeit analysieren, das Rekrutierungsinterview zu geben. Dafür informieren die Kapitel 5 bis 8 zunächst über die einzelnen einbezogenen Komplexe von Determinanten. Da es in der Natur der Sache liegt, dass die jeweiligen Variablengruppen ihrer inhaltlichen Art nach immer nur von Teilmengen der Stichprobe vorliegen können, entsteht unweigerlich ein spezielles Muster fehlender Werte. Kapitel 10 beschreibt daher, wie vor diesem Hintergrund die Einzelinformationen in ein Gesamtmodell zur Schätzung der anfänglichen Antwortbereitschaft eingebunden werden können. Zuvor informiert Kapitel 9 noch über Interviewereffekte und wie sie in vorliegendem Kontext erklärt werden können.

Kapitel 11 beschreibt unser Gesamtmodell zur Erklärung der Wiederbefragungsbereitschaft und der anschließenden Panelteilnahme. Darin enthalten sind als erklärende Teilmenge auch Integrationsindikatoren. Um deren Einfluss allerdings nicht nur im Zusammenhang mit allen anderen Einflussfaktoren, sondern auch losgelöst davon beurteilen zu können, bündelt Kapitel 12 diese Integrationsanalysen. Kapitel 13 informiert abschließend über die Ergebnisse unserer Experimente zu möglichen Befragungseffekten.

2. Umfrageteilnahme: Theorie und Forschungsstand

2.1 Allgemeine Theorien zur Umfrageteilnahme

Wer nicht an Umfragen teilnimmt, gilt als Nonrespondent. Welche Gründe gibt es dafür? Im Folgenden werden geltende Theorien dazu erläutert, die in den späteren Analysen Anwendung finden. Weiterhin wird auf den aktuellen Forschungsstand eingegangen.

Dillman et al. (2002) geben einen Überblick über die verschiedenen Ausfallgründe für die Modi postalisch, online, telefonisch und persönlich-mündlich.

Tabelle 2.1.1: Ausfallgründe – Modifizierte und übersetzte Darstellung aus Dillman et al. (2002: 6)

Umfragemodus	Grund für Non-response	Beispiele
Postalisch	Nicht-Kontaktierung	Keine Zustellung möglich
	Verweigerung	Anfrage gelesen, aber ignoriert
	Unfähigkeit	Analphabetismus
Online	Nicht-Kontaktierung	E-Mail nicht bekommen
	Verweigerung	E-Mail bekommen, aber kein Interview bzw. Interview auf erster Seite verlassen
	Unfähigkeit	Computer kann Befragung nicht downloaden
Telefonisch	Nicht-Kontaktierung	Niemand geht an das Telefon
	Verweigerung	Kontaktierte Person legt auf
	Unfähigkeit	Person ist taub, Sprach-/Verständigungsprobleme
Persönlich-mündlich	Nicht-Kontaktierung	Niemand öffnet die Tür
	Verweigerung	Kontaktierte Person verweigert Anfrage
	Unfähigkeit	Sprach-/Verständigungsprobleme

Unterschieden werden muss zwischen Nicht-Kontaktierbarkeit, Verweigerung und Unfähigkeit im Sinne von beispielsweise technischen Problemen.

Auch Groves/Couper definieren die Antwortneigung als Produkt dreier Wahrscheinlichkeiten: 1. Wahrscheinlichkeit, den Zielhaushalt/Zielperson zu lokalisieren, 2. Kontaktwahrscheinlichkeit und 3. Kooperationswahrscheinlichkeit (Groves/Couper 1998). Die Wahrscheinlichkeit, den Haushalt bzw. die Zielperson zu lokalisieren, bezieht sich meist auf Panelstudien, bei denen ein und derselbe Haushalt bzw. Zielperson wiederholt kontaktiert wird. Kontakt- und Kooperationswahrscheinlichkeiten treffen hingegen auch bei Querschnittstudien in allen Modi zu.

Trends

Berichte über gesunkene Antwortraten gibt es schon lange. Einen der frühen Berichte liefert Steeh (1981). Steeh untersucht die Antwortraten zweier Surveys in den USA (Survey of Consumer Attitudes; National Election Studies) von 1952 bis 1979 und findet einen deutlichen Anstieg von Nonresponse über die Zeit, welcher in städtischen Gebieten deutlich ausgeprägter ist. Curtin/Presser/Singer (2005: 91) finden anhand des University of Michigan's Survey of Consumer Attitudes heraus, dass bei Verweigerungen von 1979 bis 2003 ein Anstieg von 19 Prozent auf 27 Prozent zu verzeichnen ist. Ein Trend für die Nichterreichbarkeit ist in der Studie nicht festzustellen.

De Heer (1999) sowie de Leeuw und de Heer (2002) zeigen ebenfalls anhand internationaler Vergleichsstudien einen Anstieg von Nonresponse auf. De Heer (1999) vergleicht die Antwortraten von 16 Ländern aus dem Labour Force Survey. Da es sich bei dieser Studie um eine Befragung der offiziellen Statistik handelt, liegen die Antwortraten mit über 80 Prozent insgesamt sehr hoch, dennoch sind sinkende Tendenzen in den Antwortraten von 1983 bis 1997 zu erkennen. Auch de Leeuw und de Heer (2002) finden steigende Nonresponse-Raten bei der Analyse offizieller Statistiken, die zudem über die Länder variieren. Battaglia et al. (2008: 535) beobachten einen Rückgang der Antwortraten im National Immunization Survey von 1997 mit 84,6 Prozent bis 2003 mit 69,8 Prozent. Wübbenhorst/Wildner (2008) analysieren die Antwortraten des ALLBUS von 1982 bis 2004. 1982 lag die Antwortrate bei 69,7 Prozent und 2004 nur noch bei 45,7 Prozent.

Gründe für Nicht-Kontaktierbarkeit

Steigende Raten der Nicht-Kontaktierbarkeit werden in der Theorie durch sozio-politische Veränderungen erklärt. Tucker/Lepkowski (2008: 12) führen dazu beispielhaft für die USA aus, dass Personen weniger zu Hause sind, da der Anteil von Hausfrauen in der Gesellschaft sinkt. Auch Porst (1996: 21) führt steigende Ra-

ten der Nichterreichbarkeit auf gesellschaftliche Veränderungen, wie vermehrte Singlehaushalte sowie auf allgemein weniger gesellschaftliche Partizipation, zurück. Um Kontaktwahrscheinlichkeiten zu erhöhen, gibt es für alle Modi bestimmte Techniken. Beispielsweise gibt es diverse Empfehlungen, in welchen Intervallen Haushalte per Telefon kontaktiert werden sollten (vgl. z.B. D'Arrigo/Durrant/Steele 2009).

Gründe für oder gegen Kooperation

Eine besondere Herausforderung stellt aber die Kooperationswahrscheinlichkeit dar. Verweigerungen werden allgemein durch schlechte Einstellungen zu Umfragen, veränderte Einstellungen zum Datenschutz und zum Vertrauen allgemein, Angst vor Kriminalität (Beispiel Haustürverkäufe), ein allgemein schlechtes Umfrageklima oder schlicht Desinteresse erklärt (vgl. Engel et al. 2004: 1; Porst 1996: 21; Tucker/Lepkowski 2008: 12; Wübbenhorst/Wildner 2008: 20).

Soziodemographie als erfolgloser Erklärungsversuch

Oft wurde versucht, soziodemographische Faktoren zur Erklärung von Verweigerungen heranzuziehen. Viele Studien haben aber gezeigt, dass der Erklärungsgehalt solcher Faktoren meist gering ist. Da soziodemographische Verteilungen oft vorhanden sind (z.B. Mikrozensus), könnten Verzerrungen durch Gewichtungen entgegen gesteuert werden. Einzig über einen Mittelschichtbias sind sich die Forscher einig (vgl. Esser 1973; Goyder 1987; Krebs 1995). Demnach seien Personen aus der Mittelschicht in Umfragen überrepräsentiert.

Nach Schnell (1997: 138) gibt es aber nur geringe Zusammenhänge zwischen soziodemographischen und inhaltlichen Variablen, weshalb auf Basis von soziodemographischen Variablen keine Aussagen über den Grad der Verzerrung getroffen werden können.

Rational-Choice-Theorie & Theory of planned behavior

Populär sind hingegen Rational-Choice-Theorien zur Erklärung der Umfrageteilnahme. Teilnahme und Verweigerung sind dabei Entscheidungsalternativen, die nach einer reinen Kosten-Nutzen-Rechnung abgewogen werden. Als Opportunitätskosten stehen dabei die Zeit, die kognitive Last der Befragung oder die Befürchtung von Konsequenzen gegenüber dem Nutzen, der sich aus dem Interesse am Thema, Neugierde oder dem Gefühl, etwas für die Gesellschaft zu tun, ergibt (vgl. Engel et al. 2004: 55 ff.). Rational-Choice-Theorien können als zweckrationale Theorien nach Max Weber verstanden werden, bei denen die Kosten-Nutzen-Abwägung rein subjektiv erfolgt (vgl. ebd. 57).

Als Spezialfall der Rational-Choice-Theorie gilt die Theorie der überlegten Handlung (Theory of reasoned action) nach Fishbein/Ajzen (1975) bzw. des geplanten Verhaltens (Theory of planned behavior) nach Ajzen (1991). In diesen Theorien gilt die Verhaltensabsicht, die durch Einstellungen, soziale Normen und in der Theorie des geplanten Verhaltens zusätzlich durch Kontrollüberzeugungen bedingt ist, als wichtigster Erklärungsfaktor für das Verhalten. Diese Theorien dienten schon in einigen wenigen Studien recht erfolgreich der Vorhersage der Umfrageteilnahme (vgl. Bosnjak 2002; Hox/de Leeuw/Vorst 1995; Leopold 2004; Vehre 2011).

Wertrationalität & Austauschtheorie

Engel et al. (2004: 57) betonen zusätzlich die wertrationale Komponente. Dabei wird Verhalten durch Überzeugungen erklärt und nicht durch reine Kosten-Nutzen-Abwägungen. Eine Umfrageteilnahme wäre nach diesem Denkansatz trotz Opportunitätskosten vorstellbar. So finden Porst/von Briel (1995: 2) heraus, dass beim ZUMA-Methodenpanel ein Drittel der Befragten aus altruistischen Gründen teilgenommen hat.

Im Grunde stammen Hypothesen über Umfrageteilnahme aus drei theoretischen Bezugsrahmen: Rational Choice, sozialer Austausch und soziale Integration. Max Webers klassische Unterscheidung von zweck- und wertrationalem Handeln könnte helfen, in der Frage der Umfrageteilnahme eine Brücke zwischen diesen Bezugsrahmen zu schlagen. Die Teilnahme an einer Umfrage kann zweckrational motiviert sein und folgt dann im Kern einem Modell, bei dem Kosten- und Nutzenerwägungen, also antizipierte Handlungskonsequenzen bestimmen, welches Handlungsziel (Kooperation oder Verweigerung) verfolgt wird. Die Vorstellung, dass die Teilnahme an einem Interview interessant sein bzw. Spaß bereiten könnte, wäre sicherlich ein Beispiel dafür. Warum aber wird an Umfragen teilgenommen, wenn sich die angefragte Person selbst gar keinen Vorteil welcher Art auch immer für sich davon verspricht? Selbst bei Einsatz von im Vorhinein ausgezahlten monetären Anreizen wäre es, zweckrational betrachtet, eigentlich für jede Person rationaler, ein Umfragegesuch zurückzuweisen, wenn sie sich selbst gar keinen Vorteil von einer Teilnahme verspricht: der Anreiz ist ihr bereits ausgezahlt worden, ohne dass sich daraus eine Gegenleistung einfordern ließe. Zugleich erfordert die Teilnahme am Interview in jedem Fall zusätzlich zu investierende Zeit und mag auch Kosten anderer Art erzeugen. Eine einfache Erklärung dafür liegt dann in der Erkenntnis, dass das Verhalten nicht zweck-, sondern wertrational motiviert ist und sich dabei an Werten wie Gegenseitigkeit¹,

1 Das Gegenseitigkeits- bzw. Reziprozitätsprinzip kann völlig im Eigeninteresse verfolgt werden. »Tit for tat« oder »wie Du mir, so ich Dir« ist vielleicht sogar der Prototyp zweckrationalen

Verantwortung oder Hilfsbereitschaft und dergleichen orientiert, und dass es evtl. vom Grad der Integration einer Person in die umgebende Gesellschaft (Nachbarschaft, Gemeinde, Vereine und Verbände, etc.) abhängt, ob solche Werte handlungs- bzw. entscheidungsrelevant werden. Eine solche Überlegung läge auf der Linie einer Forschungsrichtung, die die Rolle von Integrationsvariablen wie »community attachment« oder »community involvement« bzw. betont,² dass Erwartungen und Verpflichtungen gegeneinander abgewogen werden. Couper/Groves (1996a) sprechen hier Bürgerpflichten an, die als Teilnahmemotiv gesehen werden können.

Leverage Saliency Theorie

Eine weitere populäre Theorie ist die Leverage Saliency Theorie nach Groves/Couper/Corning (2000). Dabei wurde ein Modell der Antwortbereitschaft konzipiert, bei dem eine Waage mit mehreren Haken, an denen unterschiedliche Gewichte hängen, im Vordergrund steht. Diese Waage kann sich entweder in die Richtung des Interviews oder in die der Verweigerung neigen. Die Gewichte prä-

Verhaltens. Das unterstellt allerdings, dass die Zukunft zählt. Das Prinzip wird so nur funktionieren, wenn Ego zu erwarten hat, in der Zukunft wieder auf Alter zu treffen, so dass Alter in der Zukunft die Gelegenheit erhält, Ego eine Verweigerung heimzuzahlen. Eine Person, die von einem Umfrageinstitut kontaktiert und um Teilnahme an einem Interview gebeten wird, hat aber nicht zu befürchten, dass dies der Anfang eines iterativen Prozesses sozialer Interaktionen ist, in deren Zuge sie wiederholt mit dem Institut konfrontiert werden wird. Warum also sollte diese Person die Haltung eines »wie Du mir, so ich Dir« einnehmen, wenn sich für sie künftig weder Kooperation positiv noch Verweigerung negativ auswirken kann; es sei denn, aus wertbezogenen Gründen, die sich nicht aus zweckrationalen Erwägungen ableiten. Das Gegenseitigkeitsprinzip wird dabei als Wert aufgefasst, den sich eine Person ebenso zu eigen machen kann wie andere Werte, sich aber nicht zu eigen machen muss. Ebenso variabel ist es, ob das Gegenseitigkeitsprinzip im Einzelfall utilitaristischen Erwägungen oder Vorstellungen entspringt, einen sozialen Austausch ausgewogen zu halten, ohne dass daran irgendeine Erwartung geknüpft wäre. Folgt man der Tailored-Design Methode von Dillman/Smyth/Christian (2009), so wäre etwa im Falle eines vorausbezahlten monetären Anreizes gerade zu vermeiden, dass sich eine Person dadurch für die Zeit bezahlt fühlt, die das Bearbeiten des Befragungs- bzw. Interviewgesuchs erfordert. Eher schon sollte dies als kleine Aufmerksamkeit für die Mühe gelten, die jemand freiwillig auf sich nimmt. Die Wirkung des Anreizes wäre dann nicht Ausdruck eines ökonomischen, sondern eines sozialen Tausches und würde so eher einem moralischen Verpflichtungsgefühl entspringen, fair zu bleiben, als einer abwägenden Entscheidung, wie man sie treffen würde, wenn es um die Einschätzung der Äquivalenz von Leistung und Gegenleistung bei einem Vertragsabschluss geht (Engel 2006: 191f.). Es ist schlichtweg eine empirische Frage, welche Werthaltungen eine Person hält und auf welcher motivationalen Grundlage das Gegenseitigkeitsmotiv darin zum Tragen kommt, ob auf der Basis eines Modells ökonomischen oder sozialen Austausches.

- 2 Zur Bedeutung von Faktoren, die wie »community attachment« oder »community involvement« unter dem Aspekt der sozialen Integration gesehen werden können, siehe z.B. Groves/Singer/Corning 2000: 306 oder Lepkowski/Couper 2002: 267.

sentieren dabei Attribute der Teilnahmeentscheidung. Je größer die Gewichte sind, desto stärker wird das Argument vom Befragten wahrgenommen. Der Abstand zwischen den einzelnen Haken und dem Hebelpunkt der Waage zeigt die Bedeutsamkeit der einzelnen Attribute. Die verwendeten Attribute beinhalten das Thema der Befragung, die Verwendung von Anreizen sowie die befragende Institution. Durch die Betonung der einzelnen Attribute durch den Interviewer können die Gewichte verschoben werden.

Dieses Bild soll die Multivariabilität des Forschungsgegenstandes und vor allem auch die möglichen kombinierten Effekte veranschaulichen. Die Theorie soll dahin führen, dass Interviewer lernen, flexibel zu reagieren und die Attribute bei der Interviewanbahnung betonen, denen die Befragten ein größeres Gewicht zuweisen (Tailoring).

Fünf Faktoren von Groves/Cialdini/Couper

Groves/Cialdini/Couper (1992) führen fünf weitere Einflussfaktoren der Umfrageteilnahme auf: Gesellschaftliche Faktoren, Eigenschaften des Umfragedesigns, Eigenschaften der befragten Person, Interviewereigenschaften und die Interviewer-Befragten-Interaktion.³

States & Traits

Ein neuerer Theorieversuch stammt von Schnauber/Daschmann (2008). Die Autoren unterscheiden *States* und *Traits*, wobei *States* situative Faktoren der Kontaktaufnahme sind und *Traits* stabile Einflussfaktoren wie Einstellungen, Erfahrungen oder Verhaltensweisen. Zur Analyse der *States* und *Traits* wird eine Verweigereranalyse durchgeführt. Dabei haben 27 Prozent aller Verweigerer einer Marktforschungsstudie den sogenannten Verweigererblock beantwortet. Es zeigt sich, dass sich Teilnehmer und Nicht-Teilnehmer hinsichtlich ihrer grundlegenden Einstellungen nicht voneinander unterscheiden (ebd.: 113). Allerdings hat die Umfrageerfahrung einen Einfluss dahingehend, dass eine Teilnahme eher zu erwarten ist, wenn die letzte Interviewerfahrung positiv war (ebd.). Bei der Analyse der Einstellungen zu Umfragen wird weiterhin zwischen generellen und altruistischen Bewertungen unterschieden. Generelle Bewertungen schließen ein, ob man Spaß an Befragungen hat und altruistische Bewertungen, ob man Befragungen sinnvoll findet (ebd.: 114). Von den generellen Bewertungen geht ein signifikanter positiver Effekt auf die Teilnahmebereitschaft aus. Im Gegensatz dazu steht die altruistische Bewertung, von der kein so klarer Effekt ausgeht. Daraus wird geschlossen, »[...] dass potenzielle Befragte sich nicht deshalb für die Teil-

3 Siehe auch die Arbeiten von Groves/Couper (Groves/Couper 1995; Couper/Groves 1996b).

nahme entscheiden, weil sie davon überzeugt sind, der Gesellschaft einen Dienst zu erweisen, sondern weil sie persönlich Interesse und Spaß daran haben« (ebd.: 114). Bezüglich der *States* zeigt sich, dass Personen, die aufgrund von Zeitargumenten verweigerten, eine positivere Einstellung zu Umfragen haben als Personen mit anderen Verweigerungsargumenten (ebd.: 116).

Da somit kein Zusammenhang mit tieferen Persönlichkeitsmerkmalen besteht, wird schließlich davon ausgegangen, dass keine großen Verzerrungen durch Selbstselektion zu erwarten sind, wobei die Generalisierbarkeit dieses Ergebnisses dadurch eingeschränkt bleibt, dass nicht alle Verweigerer Auskunft über die Gründe der Verweigerung gegeben haben (ebd.: 119). Bock/Schnapp (2009) stellen das Ergebnis von Schnauber/Daschmann in Frage, da die getroffenen Annahmen nicht ausreichend abgesichert seien und die Interpretation der Ergebnisse zu weitreichend sei.

Verwendung von Parاداتen

In der Regel ist es schwierig, die eben beschriebenen Theorien einer vollständigen empirischen Überprüfung zu unterziehen, da dafür die Informationen für Nonrespondenten und Respondenten zur Verfügung stehen müssten. Wenn Personen nicht an der Umfrage teilnehmen, dann liegen zum Beispiel über die eben erwähnten Persönlichkeitsmerkmale keine Informationen vor. Daher kommen in Nonresponse-Analysen Parاداتen immer mehr zum Einsatz. Unglücklicherweise sind Parاداتen in öffentlich zugänglichen Datensätzen nur selten vorhanden.

Für diese Art von Daten, die neben den inhaltlichen Interviewvariablen erhoben werden, gibt es neben dem Begriff Parاداتen eine Reihe von weiteren Begriffen, wie Kontaktdaten, Prozessdaten, Metadaten oder im Englischen Summary Data oder Call Record Data. Alle diese Begriffe meinen im Prinzip dasselbe: Sie beschreiben den Prozess der Datenerhebung innerhalb der Interaktion zwischen Befragtem und Interviewer (Blom/Lynn/Jäckle 2008; Durrant/D'Arrigo/Steele 2008; Kennickel/Mulrow/Scheuren 2009; Kreuter/Kohler 2009).

Couper (2000) hat den Begriff Parاداتen als globalen Begriff für die Beschreibung des Erhebungsprozesses eingeführt. Der Autor definiert Parاداتen als Hilfsvariablen, die eben diesen Prozess beschreiben in Abgrenzung zu Metadaten, die wiederum Daten als solches beschreiben.

Weiterhin gibt es inhaltliche Unterscheidungen. Einige Autoren empfehlen, zwischen Mikro- und Makroparاداتen zu unterscheiden (Safir/Black/Steinbach 2001; Scheuren 2001; Scheuren 2005). Makroparاداتen beziehen sich dabei auf globale zusammenfassende Merkmale wie die Teilnahme- oder Verweigerungsrate (Safir/Black/Steinbach 2001; Scheuren 2005). Solche Daten sind sehr gebräuchlich, da sie ein praktisches Nebenprodukt computergesteuerter Umfragen sind. Mikroparاداتen beziehen sich detaillierter auf den Kontaktprozess. Zu diesen

zählen beispielsweise Informationen über die Interviewsprache, über Kontaktversuche, Kontaktsergebnisse, Dauer und Zeit des Interviews, Informationen über Überzeugungsversuche oder im Onlinefall Browserinformationen (Kennickel/Mulrow/Scheuren 2009; Kreuter, Casas-Cardero 2010; Laflamme 2008; Sawtooth Software 2002; Scheuren 2005).

Mikroparadaten werden also teilweise durch den Interviewer erhoben, wie bei der Erfassung der Überzeugungsversuche. Dies ist äußerst zeit- und kostenintensiv und deshalb nicht sehr gebräuchlich.

Eine andere Art von Daten über Daten sind sogenannte Kontextdaten, die üblicherweise aus Stichproben den Interviewdaten zugespielt werden (Kennickel/Mulrow/Scheuren 2009: 1). Beispiele für solche Kontextdaten sind Regionalcodes oder in persönlich-mündlichen Befragungen Informationen über die Nachbarschaft.

Auch Kreuter/Kohler (2009) betonen, dass dem Kontaktprozess bei der Analyse von Nonresponse mehr Aufmerksamkeit geschenkt werden muss und nutzen in ihrer Studie die Ergebnisse der einzelnen Kontaktversuche als Nonresponse-Gewichte.

2.2 Besonderheiten der Umfrageteilnahme nach Modus

2.2.1 Teilnahme an Umfragen: online

Bei der Teilnahme an Onlinebefragungen gibt es einige Ergebnisse, die darauf hindeuten, dass diese auch unter Kontrolle der Internetnutzung (die stark mit den soziodemographischen Variablen korreliert ist) mit soziodemographischen Variablen zusammenhängt. Vehovar et al. (2002) beobachteten hier einen fast linearen Zusammenhang zwischen den Variablen, die die Internetnutzung beeinflussen und der Teilnahmebereitschaft (ebd.: 34). Ähnliche Ergebnisse, die für einen Zusammenhang zwischen der Teilnahme an Online-Surveys und soziodemographischen Variablen sprechen, finden sich bei Couper et al. (2007) und Bandilla et al. (2009). Hier ist zu vermuten, dass die Teilnahmebereitschaft bei Onlinebefragungen mit der Intensität der Nutzung zusammen hängt, die wiederum sehr wahrscheinlich durch soziodemographische Merkmale erklärbar ist (vgl. Bartsch 2011).

So gibt es Hinweise darauf, dass die Teilnahme an Onlinebefragungen mit der vorherigen Erfahrung mit dem Internet zusammenhängt (El-Menouar/Blasius 2005; Tuten/Urban/Bosnjak 2002; Faas 2003). Einen ähnlichen Befund berichten Fricker et al. (2005). Sie verglichen Internetnutzer, die sie für eine Befragung zufällig auf die beiden Modus-Bedingungen Internet und Telefon verteilt hatten. Die Personen, die an der Onlinebefragung teilnahmen, hatten ein größeres Wis-

sen über Technik als diejenigen Internetnutzer, die in der Telefonbedingung antworteten. Allerdings geben die Autoren zu bedenken, dass die Unterschiede in den Verteilungen bei den Wissensfragen auch mit dem Modus der Datenerhebung zu tun haben können: Bei einer Telefonbefragung ist es kaum möglich, etwas länger über eine Frage nachzudenken. In der Onlinebedingung wäre es theoretisch sogar möglich, eine Antwort nachzuschlagen. Es ist aber dennoch plausibel zu vermuten, dass Teilnehmer an Onlinebefragungen eine höhere Affinität zum Internet und zu technischen Entwicklungen haben. In einigen Arbeiten zur Teilnahme an Onlineumfragen wird darüber hinaus die Neugier als ein Motiv genannt, an solchen Umfragen teilzunehmen (Baur/Florian 2009).

Einen weiteren Hinweis darauf, dass die Teilnahme an Onlinebefragungen mit der Erfahrung und der Häufigkeit der Nutzung des Internets zusammen hängt, liefern die Befunde dazu, dass sich Früh- und Spätantworter in Befragungen genau in diesem Merkmal unterscheiden (Archer 2007; Gadeib 2005; Biffignandi et al. 2002). Leopold (2004) hat in seiner Arbeit mögliche Gründe für Nonresponse in Onlinebefragungen zusammengestellt. Diese sind:

1. Unfähigkeit, die Befragung zu öffnen
2. Probleme innerhalb der Befragung mit dem Navigieren
3. Probleme mit der Kompatibilität von Programmen/Browsern
4. Computerprobleme bei der Zielperson
5. Netzwerkprobleme
6. Probleme technischer Art der Befragung (Serverproblem)

Insbesondere die Gründe 1 und 2 weisen daraufhin, dass der Befragte eine gewisse Erfahrung mit dem Internet benötigt, um erfolgreich an der Befragung teilzunehmen. Hier ist auf der anderen Seite allerdings auch das Befragungsinstitut gefordert. Sowohl das Aufrufen der Befragung als auch das Navigieren innerhalb der Befragung sollte möglichst einfach und leicht verständlich erklärt sein (Bosnjak/Tuten 2001). Zusätzlich ist beim Einsatz von Bildern, Filmen und Plausibilitätschecks zu beachten, dass die Downloadzeit für den Befragten nicht zu lang werden sollte (Pfleiderer 2001), denn dadurch können Personen mit älterer Software und/oder langsamer Internetverbindung an der Teilnahme gehindert werden.

2.2.2 Teilnahme an Umfragen: telefonisch

Bei telefonischen Umfragen sind zwei Punkte von zentraler Bedeutung: 1. der Umstand, dass eine fremde Person anruft und 2. Interviewereffekte. Punkt 1 hat sich durch zahlreiche Datenmissbrauchsskandale und verstärkt auftretende Werbeanrufe unter dem Deckmantel von Befragungen zu einem enormen Problem

entwickelt. Das Vertrauen in telefonische Interviewanfragen ohne Vorabinformationen scheint stark geschwächt. Vorankündigungen sind aber nur bei speziellen Befragungen möglich, bei denen neben der Telefonnummer auch eine Adresse vorhanden ist. Bei Bevölkerungsbefragungen mit zufällig generierten Telefonnummern ist dies nicht möglich. Es ist somit der Interviewer, der den ersten Kontakt herstellt und von der Teilnahme überzeugen muss.

In persönlich-mündlichen Interviews wirkt der Interviewer durch Gestik, Mimik, Alter und Auftreten. Interviewereffekte am Telefon lassen sich in Überzeugungsargumente und stimmliche Merkmale der Interviewer einteilen. Groves/Lyberg (1988) fügen zusätzlich Effekte des Geschlechts und der Erfahrung an. Demnach zeigten weibliche Interviewer positive Effekte.

Effekte von Überzeugungsargumenten werden ausführlich bei Dijkstra/Smit (2002) untersucht. Die Autoren konnten herausfinden, dass es förderlich ist, nicht standardmäßig etwas über die Länge des Interviews zu sagen (ebd.: 126). Wird explizit vom Befragten nach der Dauer des Interviews gefragt, führen abschwächende Formulierungen, wie »es dauert nicht lange«, zu den meisten Interviews. Weiterhin konnte herausgefunden werden, dass Wiederholungen der Verweigerungsargumente durch den Interviewer eher Verweigerungen verstärken. Äußert jemand beispielsweise, kein Interesse zu haben und der Interviewer antwortet mit »Sie haben kein Interesse?« ist es für den Befragten aus psychologischer Sicht leichter, die Anfrage erneut zu verneinen, seine Meinung also zu bestätigen, anstatt sie zu revidieren (ebd.: 127). Stattdessen werden von den Autoren Ermutigungen, wie »Sie würden uns aber sehr helfen!«, empfohlen.

Ein weiterer Punkt ist das Tailoring. Dabei verwendet der Interviewer die Informationen, die der Befragte liefert (Verweigerungsgründe, Stimme, Stimmung) und passt sein Überzeugungsargument dementsprechend an (vgl. ebd.: 129). Die Studie konnte zeigen, dass direkt auf den Befragten zugeschnittene Überzeugungsstrategien mehr zum Erfolg führen als nicht direkt zugeschnittene Strategien (ebd.: 130). Die Effekte sind allerdings nur nahezu signifikant.

Stimmliche Merkmale der Interviewer

Als Persönlichkeitsmerkmale treten bei Telefonumfragen stimmliche Merkmale in den Vordergrund. Erstmals untersuchen Hüfken/Schäfer (2003) den Einfluss stimmlicher Merkmale auf die Teilnahmebereitschaft in Deutschland. Dem voraus geht eine Studie aus den USA von Oksenberg/Cannell (1988). Stimmeigenschaften werden hier als paralinguistische Charakteristika bezeichnet (ebd.: 258). Interessanterweise erzielen Interviewer, die schnell, laut und mit einer standardmäßigen Betonung sprechen, den Autoren zufolge die höchsten Antwortraten (ebd.: 267).

Hüfken/Schäfer (2003) vergleichen den Einfluss stimmlicher und persönlicher Eigenschaften der Interviewer des ALLBUS 2000 in Deutschland. Als Er-

gebnis wird festgestellt, dass persönliche Eigenschaften einen weitaus größeren Einfluss haben als stimmliche Merkmale. Freundlichen Interviewern, die eine fröhliche Atmosphäre schaffen und die Befragten persönlich ansprechen, werden dabei die größten Chancen eingeräumt (ebd.: 336). Ergebnisse der Stimmanalyse zeigen, dass frei gesprochene Einleitungen mit lauterer Stimme, geringerer Variation der Stimmhöhe und korrekter Betonung eher zum Interview führen (ebd.). Weiterhin wirkt es sich im Gegensatz zu den Ergebnissen aus den USA von Oksenberg/Cannell (1988) positiv aus, langsamer zu sprechen (ebd.).

2.2.3 Teilnahme an Umfragen über das Mobiltelefon

Die Befragung über Mobiltelefone ist eine recht neue Methode. Die Teilnahmebereitschaft an solchen Umfragen wurde von Schneiderat/Schlinzig (2009) eingehend untersucht. Die Autoren fanden in der CELLA-Studie heraus, dass die Teilnahmebereitschaft bei der Mobilbefragung (33,7 %, kalter Kontakt, keine Vorankündigung) höher war als bei der durchgeführten Befragung über das Festnetz (26,5 %, kalter Kontakt). Die Autoren begründen dies dadurch, dass solche Befragungen noch neu sind und demnach nicht mit dem schlechten Image von normalen Telefonbefragungen durch Werbeanrufe verbunden werden (ebd.: 97). Im Bereich von Umfragen über das Festnetz sprechen die Autoren von einer Sättigung (ebd.). Bei Mobilfunkumfragen besteht zudem der Vorteil, die Umfragen per SMS ankündigen zu können. Im Rahmen der CELLA-Studie wurde dies mit einem Hinweis auf eine Homepage mit weiteren Informationen realisiert. Die Teilnehmerate konnte durch die Ankündigung per SMS noch erhöht werden (ebd.). Vertrauensbildend wirkt hier auch die Übermittlung der Absendernummer.

Weiterhin gilt für Mobilbefragungen genau wie bei Befragungen über das Festnetz, dass mit Interviewereffekten zu rechnen ist.

2.3 Antwortneigung

Um die Antwortneigung von Personen zu schätzen, sind relevante Informationen über Umfrageteilnehmer und Nichtteilnehmer (Ausfälle) erforderlich. Umfragedaten liegen aber naturgemäß nur von Leuten vor, die an Umfragen teilnehmen. Eine Möglichkeit, Informationen über Nichtteilnehmer in die Modellierung der Antwortneigung einzubeziehen, liegt in der Verwendung von externen Registerdaten, wie sie im Falle von Einwohnermeldeamtstichproben verfügbar sind. Werden aus Melderegistern Stichproben gezogen und auszuschöpfen versucht, so

fallen dabei in aller Regel Ausfälle von Personen an, die nicht erreicht werden oder nicht kooperieren möchten. Umfragedaten liegen dann nur von Umfrageteilnehmern vor. Die im Zuge der Stichprobenziehung anfallenden Registerdaten liegen allerdings nach wie vor für beide Teilgruppen vor, und zwar von Teilnehmern wie Nichtteilnehmern an der Umfrage. Gleiches gilt für Regionalinformationen, die den Wohnort bzw. die Umgebung des Wohnsitzes charakterisieren und über die Kenntnis der Adresse der zu befragenden Personen zu den Umfrage- und Paradata hinzu gespielt werden können.

Die Verwendung externer Register- und Regionalinformationen scheint daher eine praktikable Möglichkeit zu eröffnen, die Neigung (Wahrscheinlichkeit) von Personen zu schätzen, auf Umfragegesuche positiv zu antworten, das heißt: an der Umfrage teilzunehmen.

Im Prinzip ist diese Einschätzung zutreffend. Einer Modellierung über externe Referenzdaten sind nach vorliegenden Erfahrungen gleichwohl enge Grenzen gesetzt. Diese werden zum einen dadurch gezogen, dass in den Registern von beispielsweise Einwohnermeldeämtern nur wenige inhaltliche Informationen enthalten sind. Engel/Schnabel/Simonson (2006) untersuchten die Nützlichkeit dieses Ansatzes anhand von drei kommunalen Befragungsstudien und fanden heraus, dass sich die Wahrscheinlichkeit einer Umfrageteilnahme nur zu einem geringen bis vernachlässigbaren Anteil aus den verfügbaren soziodemografischen Variablen (Geschlecht und Alter) und Charakteristika der Nachbarschaften und Stadtteile, in denen die Personen leben, erklären ließen. Die erklärten Varianzen lagen zwischen 1,1 und 3,2 Prozent (nach den Pseudo- R^2 der durchgeführten logistischen Regressionen). Selbst wenn jede Stichprobenperson in Ergänzung zur individuellen Soziodemographie durch die Zugehörigkeit zur umgebenden Nachbarschaft selbst (und nicht durch Charakteristika dieser Nachbarschaft) klassifiziert wurde, ließ sich der Anteil erklärter Varianz nur auf 6,1 bis 9,7 Prozent steigern.

Die Möglichkeiten, ausfallbedingte Verzerrungen von Verteilungen der inhaltlichen Zielvariablen einer Umfrage durch Bezug auf externe Referenzverteilungen zu korrigieren, werden allerdings noch durch einen zweiten Faktor stark begrenzt, und zwar durch die Höhe der Korrelation zwischen diesen Zielvariablen einer Umfrage und den externen Referenzvariablen. Diese Korrelation ist stets zu beachten (Bethlehem 2002: 287), wenn es darum geht, Verzerrungen zu kompensieren, die inhaltlichen Zielvariablen einer Umfrage zuzurechnen sind und nicht den externen Hintergrundvariablen, auf die sich eine Modellierung und kompensatorische Gewichtung von Ausfällen gegebenenfalls stützt (Dillman et al. 2002: 19).

Sofern also erklärungskräftige Register- und Regionalinformationen nur sehr eingeschränkt vorliegen und »frame data or external population data [...] at best modestly correlated with the survey variables« sind (Lynn 2003: 239), dann kann

der Bezug auf externe Referenzdaten im Allgemeinen nur von begrenzter Nützlichkeit sein, wenn es darum geht, ausfallbedingte Verzerrungen in den Verteilungen der inhaltlichen Zielvariablen einer Umfrage zu kompensieren.

Es ist daher sinnvoll, die Analyseperspektive um mögliche Alternativen zu erweitern. Als erfolgversprechend erschienen uns dazu insbesondere Versuche, relevante Informationen aus dem Umfrageprozess selbst zu gewinnen, wie etwa in den Ansätzen von Czajka et al. (1992), Kessler/Little/Groves (1995), Curtin/Presser/Singer (2002), Lynn et al. (2002), Lynn (2003) und Voogt/Saris (2005).

In einer Studie zur Verkehrsmittelwahl (Engel/Hansen 2001; Engel/Pötschke 2003), die vor dem Hintergrund der Klimaschutzfrage durchgeführt wurde, hatten wir beispielsweise einen ähnlichen Ansatz wie Lynns PEDAKSI Methodologie verfolgt, um die Antwortbereitschaft zögerlicher Personen über das Angebot einer kürzeren Kernversion des vollen Interviews zu verbessern (Engel/Schnabel/Simonson 2006). Dadurch ließ sich die Antwortrate um den Faktor 1,3 erhöhen. Bedeutsamer als dieses Ergebnis war in diesem Zusammenhang aber ein anderes. Hätten wir die Möglichkeit des Kerninterviews nicht unterbreitet, so wären die Teilnehmer daran zu Nichtteilnehmern (Ausfälle) geworden. Wir hätten deren Antworten also nicht verfügbar gehabt. Nun zeigte sich aber, dass die Teilnehmer des Kerninterviews in Bezug auf die zentrale abhängige Variable der Studie inhaltlich anders geantwortet haben als die Teilnehmer der Vollinterviews. Es ging um die Frage, ob für die täglichen Wege zur Arbeit oder Ausbildungsstätte das Auto oder öffentliche Verkehrsmittel genutzt werden. Von den Teilnehmern des Vollinterviews gaben 56 Prozent an, dafür das Auto zu nutzen, während es unter den Teilnehmern des Kerninterviews 78 Prozent waren. Ohne das Angebot einer Kernversion hätte sich die Schätzung der Autonutzung in der Population nur auf die 56 Prozent stützen können. Das aber wäre eine Unterschätzung des wahren Werts gewesen, da die Frage der Teilnahme/Nichtteilnahme an dieser Studie offenbar auch eine Funktion des Themas der Befragung (Verkehrsmittelwahl) war und einige daran offenbar zunächst nicht mitwirken wollten, um gewissermaßen nicht als Klimasünder dazustehen. Da nun aber auch Angaben von Personen vorliegen, die unter normalen Bedingungen nicht an der Studie teilgenommen hätten, besteht nun die Möglichkeit, die Schätzung des Ausmaßes der Autonutzung zu korrigieren. Darauf, dass wir mit ausfallbedingten Verzerrungen zu rechnen haben, wenn die Umfragethematik einen Faktor darstellt, der die Teilnahmeentscheidung selbst berührt, und zwar mit »nonignorability implications for key statistics«, weisen auch Groves/Presser/Dipko (2004: 2) hin.

Individuelle Antwortneigungen können nur auf der Basis von Modellen geschätzt werden. Eine in diesem Zusammenhang diskutierte Frage ist, ob ein ein-dimensionaler Raum angenommen werden kann, der sich vom Pol der leicht erreichbaren und zugleich kooperativen Personen zum Pol derjenigen erstreckt, die trotz aller etwaigen Arrangements von Umfragedesigns und trotz aller Anstren-

gungen während der Feldphase Nichtteilnehmer, also Ausfälle, bleiben.⁴ Die Annahme eines solchen Kontinuums hätte den Vorteil, die Linie von unmittelbar befragungsbereiten Personen hin zu Personen, die nur mit steigendem Aufwand zu einer Umfrageteilnahme bewegt werden können, über den Pol hinausgehend zu denjenigen Personen zu verlängern, die auch bei größten Anstrengungen nicht an einer Umfrage teilnehmen würden. Ob aber eine solche Linie zunehmender Zurückhaltung angenommen und zur Extrapolation auf extremere Fälle hin eingesetzt werden kann, ist fraglich. Ebenso fraglich ist entsprechend, ob mit zunehmender Distanz vom Pol unmittelbarer Kooperationsbereitschaft die Ähnlichkeit zu denjenigen zunimmt, die auch bei größten Anstrengungen Nichtteilnehmer an Umfragen bleiben würden.

Vorliegende Studie folgt dieser Diskussion, in dem sie bei der Schätzung von Antwortneigungen zum einen Faktoren berücksichtigt, die Einfluss auf die Kontaktwahrscheinlichkeit ausüben, und zum anderen Faktoren mit möglichem Einfluss auf die Kooperationswahrscheinlichkeit. Was die Modellierung der Schwierigkeit anbelangt, den erforderlichen Kontakt zu einer Zielperson herzustellen, wird ein über die Anzahl der Kontaktversuche hinausgehender typologischer Ansatz verfolgt (siehe unten Kapitel 5 bis 8 und Kapitel 10).

2.4 Einfluss von Interviewern

In persönlich-mündlichen und telefonischen Interviews stellt der Interviewer eine der wichtigsten Hilfen dar, Antwortraten zu verbessern (Campanelli/Sturgis/Purdon 1997; Groves/Cialdini/Couper 1992; Morton-Williams 1993). Es konnte gezeigt werden, dass es zwischen Interviewern zu beträchtlicher Variation in Antwortraten kommt (Lyberg/Lyberg 1991; Lyberg/Dean 1992). Verschiedene Studien haben daher den Einfluss des Interviewers auf Antwortraten eingehender untersucht und herausgefunden, dass nur wenig Evidenz dafür besteht, dass Eigenschaften von Interviewern wie etwa Alter und Geschlecht Antwortraten beeinflussen. Jedoch liegt Evidenz dafür vor, dass die Erfahrung von Interviewern und Einstellungen von Interviewern die Antwortwahrscheinlichkeit beeinflussen (Groves/Fultz 1985; Couper/Groves 1992; de Leeuw/Hox 1996; de Leeuw et al. 1997).

Beachtet wurde zudem der Einfluss von Interviewern auf die Datenqualität. Es müssen nicht dieselben Eigenschaften von Interviewern sein, die dazu beitra-

4 Stoop (2005) diskutiert dieses und andere Modelle. Siehe auch Lin/Schaeffer (1995) und Heerwegh/Abts/Loosveldt (2007). Bzgl. der Konsequenzen besonderer Intervieweranstrengungen auf ausfallbedingte Verzerrungen, siehe Lynn et al. (2002) und Keeter et al. (2000).

gen, hohe Antwortraten zu erzielen, wie Eigenschaften, die helfen, qualitativ hochwertige Umfragedaten zu erzielen. In einer neueren Studie hat beispielsweise Stoop (2005) herausgefunden, dass Versuche, zögerliche Personen von einer Umfrageteilnahme zu überzeugen, nicht notwendigerweise auch die Qualität des resultierenden Samples verbessert.

Hox/de Leeuw (2002) verwendeten einen standardisierten Fragebogen in einer international vergleichenden Studie von Interviewereffekten auf Nonresponse. Sie erhoben Daten von 32 Umfragen für 9 Länder. Sie fanden u.a. geringe Effekte von Einstellungen von Interviewern heraus. Außerdem fanden sie beträchtliche Unterschiede zwischen verschiedenen Ländern und Studien heraus, die nur teilweise durch Eigenschaften, Einstellungen und Verhaltensweisen von Interviewern erklärt werden konnten. Gestützt auf eine psychometrische Analyse der Daten aus der Studie von Hox/de Leeuw (2002) entwickelten die beiden Autoren einen standardisierten Interviewer-Fragebogen (IQUEST). Dieser wurde beispielsweise in sieben Ländern des European Social Survey 2002 eingesetzt, ebenso in vorliegender Studie.

Die Analyse von Hox/de Leeuw (2002) unterstreicht einmal mehr die besondere Bedeutung von Einstellungen von Interviewern als Prädiktor-Variablen für interviewerspezifische Antwortraten. Die Analyse enthüllt die entscheidende Rolle, welche die Sichtweise spielt, dass zögerliche Personen überredet werden können und überredet werden sollten. Es ist diese Überzeugungsorientierung, die sich im Mehrebenenmodell von Hox/de Leeuw als stärkster, die Antwortrate erhöhender Prädiktor erweist, mit schwächeren gegenläufigen Effekten von Einstellungen, die die Freiwilligkeit der Teilnahme und den Respekt vor Verweigerung und Schutz der Privatsphäre betonen. Es scheint daher so zu sein, dass der Wirkungsweise solcher Einstellungen der einfache Mechanismus einer sich selbst-erfüllenden Prophezeiung zugrunde liegt. Wir greifen diesen Punkt in unserer Analyse erneut auf (Kapitel 9).

Geeignete Interviewer-Taktiken sind Voraussetzung dafür, die Initialsituation zwischen Interviewer und Zielhaushalt (Kontakt- bzw. Zielperson) in geeigneter Weise situationspezifisch zuschneiden zu können. Dies auch vor dem Hintergrund, dass von solchen Taktiken neben den intendierten auch unbeabsichtigte kontraproduktive Handlungseffekte ausgehen können. Die Wirkung von Interviewertaktiken und mögliche »tailoring effects« stellen daher auch wichtige Themen und Forschungsfelder dar. Einige Studien haben sich diesbezüglich mit »tailoring effects« befasst, darunter die Studien von Morton-Williams (1993), Groves/Couper (1996), de Leeuw (1999) und Groves/Singer/Corning (2000).

Groves/Couper (1996) untersuchten unter anderem, wie die Interaktion zwischen Haushalt und Interviewer das Ergebnis des Kontaktes (Interview oder Terminabsprache vs. Verweigerung) beeinflusst. Nach dem Tailoring-Konzept sollten erfolgreiche Interviewer ihr Verhalten in Reaktion auf Hinweise ändern, die

vom Gesprächspartner, also der betreffenden Person aus dem Zielhaushalt, kommen. Groves/Couper (1996) waren in der Lage, einen schwachen Indikator von »between-contact tailoring« (Veränderung des Verhaltens des Interviewers in aufeinanderfolgenden Kontakten) in ihre Studie einzubeziehen, und fanden annähernd signifikante bis signifikante Effekte in die erwartete Richtung. Werden Schätzungen von Tailoring-Effekten über sukzessive Kontakte hinweg miteinander verglichen, so unterstützt die Analyse »the expectation that the greatest opportunities for tailoring arise in early contacts« (79).

Ergänzend zum Tailoring-Aspekt beleuchtet die Studie von Groves/Couper (1996) noch einen anderen Aspekt. Und zwar, dass das Verhalten des Zielhaushalts bei einem früheren Kontakt informativ für die Erfolgsaussichten bei einem nachfolgenden Kontakt sein können. Das Verhalten des involvierten Haushaltsmitglieds wurde in drei Kategorien eingeteilt, und zwar in »negative statements«, »time-delay statements« und »questions«, wobei sich diese Kategorien vorhersage-relevant für das abschließende Ergebnis zeigten (Nur »Fragen« hatten einen positiven Effekt). Darüberhinaus können solche Statements mit substantiellen Unterschieden in der Datenqualität (Missing-Data Raten) und den Antworten verbunden sein, wie Couper (1997) zeigt: wo »too busy« Statements nur wenig systematische Variation erzeugen, erweisen sich »not interested« Statements als in beiden Hinsichten einflussreich. Couper (1997: 334f.) fasst seine Analyse so zusammen, dass Personen »who say that they are not interested in response to the interviewer's request for an interview (1) are less likely to agree to participate in the survey, (2) are less likely to provide »substantively meaningful« data (i.e., more likely to produce »missing« data) when they do so, and (3) differ in their substantive responses from those who do not express such reactions to the survey request at the outset.« Das Ergebnis legt nahe anzunehmen, dass die Statements, die seitens involvierter Haushaltsmitglieder während des Initialkontakts abgegeben werden, nicht nur für die Problematik systematischer Ausfallfehler, sondern auch für die Problematik systematischer Messfehler von Bedeutung sein können.

Groves/Singer/Corning (2000) unternahmen einen illustrativen Test ihres theoretischen Bezugsrahmens. Sie führten eine logistische Regressionsanalyse durch, um die Teilnahme in einer künftigen Umfrage vorherzusagen, und fanden empirische Evidenz heraus, wie sie von ihnen erwartet worden war: Neben positiven Haupteffekten wurde eine negative Wechselwirkung zwischen »cash incentives« und »community involvement« herausgefunden. Diese Wechselwirkung liegt auf der Linie der Theorie der Autoren und bedeutet, dass sich die Wirkung eines monetären Anreizes verringert, wenn das Engagement für die Gemeinde hoch ist. Dementsprechend ist Nichts gewonnen, wenn Versuche, eine Person zur Teilnahme an einer Umfrage zu motivieren, das falsche Motiv adressieren. Der skizzierte empirische Befund spricht sicherlich für Intervieweröffnungsstrategien, die flexibel auf situationsbezogene Hinweisreize eingehen.

De Leeuw (1999) präsentiert Evidenz über den Einfluss von Taktiken von Interviewern, Nonresponse abzuwenden, auf interviewer-spezifische Antwortraten. Als empirische Grundlage der Analyse dienen 40 Statements von 22 erfahrenen CAPI-Interviewerinnen aus einer Fokusgruppenanalyse, bei der nach der Effektivität von Taktiken gefragt war, in einer Umfrage Kooperation zu erzielen. Diese Statements wurden einer Clusteranalyse unterzogen. Es resultierten »Effektivitätscluster«, von denen »tailoring introduction«, »competent introduction« und »respondent centered« die drei höchsten subjektiven Effektivitätseinschätzungen erhielten.⁵

De Leeuw nutzte die so erzielten Cluster-Scores in einem nachfolgenden Schritt, um die Antwortraten vorherzusagen, die die beteiligten Interviewer in einem *Labour Force Survey* erzielt hatten. Dabei zeigten drei Cluster signifikante Effekte (positive Effekte: »project a positive image« und »personality«; negativer Effekt: »competent introduction«). Zwei weitere Cluster erzielten annähernd signifikante Effekte (positive Effekte: »effective public relations« und »tailor introduction«). De Leeuw (1999: 37) fasste ihre Ergebnisse wie folgt zusammen: »Experienced interviewer who emphasize the importance of personality and a positive image, are more successful than those who do not. The successful interviewers also think that tailoring and public relations are of some importance. Relatively less successful interviewers emphasize more the importance of a competent introduction.«

Die befragten Interviewerinnen der Studie von de Leeuw (1999) unterstreichen unter anderem die Bedeutung, die das Image der Einrichtung aufweist, in deren Namen sie die Interviews durchführen. Damit verweisen sie indirekt auf einen Faktor, auf den wir auch an anderer Stelle dieser Monographie zu sprechen kommen (Kapitel 4.3), und zwar auf das Image, das sich die Bevölkerung von Umfrageforschung als einer gesellschaftlichen Institution macht, und deren Beziehung zu einer anderen Institution, der (Sozial-) Wissenschaft. Die Relevanz

5 Kursiv gesetzte Statements wurden durch die Interviewer hoch eingeschätzt; Clusternamen im Fettdruck. **Tailoring introduction:** *grasping the doorstep situation; adapt introduction; adapt your introduction (modify according to social and cultural class); do not follow set of fixed rules, adapt to situation;* **competent introduction:** *mention Statistics Netherlands almost at once; start by identifying yourself;* mention introductory letter; be flexible in making appointments for interviews; know a short introduction by heart; ask if this is the right time (make an appointment if not, do not pressure); **respondent centered:** be honest (e.g. about durations, questions); react flexibly to the situation; respect the respondent; dress appropriately for neighborhood; **project a positive image:** *be friendly;* remain friendly; raise trust; be likeable; *project self-confidence;* give a relaxed impression; *be/remain yourself;* project that *you* enjoy doing your job; use intuition, feeling and humor in approach; **personality:** have perseverance; *have self-confidence (start every interaction from the perspective that the person you contact will be cooperative);* be preponderant; **public relations:** *Statistics Netherlands should pay more attention to public relations and positive image; refer to topics in the news or in society that are of importance at the moment and can be connected with the topic of the survey.*