

Online-Befragung

Methodische Grundlagen, Problemfelder,
praktische Durchführung.

Wolfgang Eichhorn



Copyright © 2004 Wolfgang Eichhorn

Rev. 1.0, München 2004

E-Mail: cc@wolfgang-eichhorn.com

Web: <http://wolfgang-eichhorn.com>

Dieses Werk kann durch jedermann gemäß den Bestimmungen der Lizenz für die freie Nutzung unveränderter Inhalte genutzt werden.

Die Lizenzbedingungen können unter <http://www.uvm.nrw.de/opencontent> abgerufen oder bei der Geschäftsstelle des Kompetenznetzwerkes Universitätsverbund MultiMedia NRW, Universitätsstraße 11, D-58097 Hagen, schriftlich angefordert werden.

Literaturangabe:

Eichhorn, Wolfgang: Online-Befragung. Methodische Grundlagen, Problemfelder, praktische Durchführung. Online-Publikation, München 2004, Rev 1.0.

URI: <http://wolfgang-eichhorn.com/cc/onlinebefragung-rev1.0.pdf>

1.	Anatomie der Befragung.....	1
1.1.	Definition.....	1
1.2.	Befragung als Abfolge methodischer Schritte	1
2.	Frageformulierung und die Messung von Antwortverhalten	3
2.1.	Frageformulierung	4
2.1.1.	Anforderungen an die Frage	4
2.1.2.	Frageformulierung: CASI	7
2.2.	Antworten und ihre Messung.....	7
2.2.1.	Messung	7
2.2.2.	Das Universum möglicher Antworten: Skalentypen	8
2.2.3.	Antworterhebung: CASI	11
2.3.	Itemlisten und Skalierungsverfahren	14
2.3.1.	Itemlisten.....	14
2.3.2.	Skalierung	14
3.	Fragebogenkonstruktion	18
3.1.	Makrostruktur.....	18
3.1.1.	Einleitung.....	18
3.1.2.	Verabschiedung.....	21
3.1.3.	Hauptteil	21
3.2.	Mikrostruktur	22
3.2.1.	Struktur von Fragesequenzen	22
3.2.2.	Fehlerquellen	23
3.2.3.	Filter	24
3.2.4.	Eingabe- und Konsistenzchecks.....	26
3.3.	Fragebogaufbau in schriftlichen Befragungen	27
3.3.1.	Fragebogaufteilung: CASI	27
4.	Stichprobenziehung	30
4.1.	Zufallsstichprobe.....	30
4.2.	Mehrstufige Auswahlverfahren	30
4.2.1.	Geschichtete Stichproben.....	30
4.2.2.	Clusterstichproben.....	31
4.2.3.	Haushalts- und Personenstichproben	31
4.3.	„Aktives“ und „passives“ Sampling.....	32
4.3.1.	Der mehrphasige Charakter einer Stichprobenziehung.....	32
4.3.2.	E-Mail-Stichproben	32
4.3.3.	Alternativen der aktiven Stichprobenziehung und Kontaktaufnahme.....	34
4.3.4.	Anwendbarkeit mehrstufiger Verfahren für die Online-Befragung.....	36
4.3.5.	Passive Verfahren	36
4.3.6.	Vollerhebungen.....	38
4.3.7.	Online-Panels	39
4.3.8.	Die Online-Population	39
4.4.	Problemfelder bei der Stichprobenziehung in der Online-Forschung.....	40
4.4.1.	Das Stichproben-Dilemma der Online-Forschung.....	40
4.4.2.	Selbstselektion bei der Stichprobenziehung.....	41
4.4.3.	Incentives als Mittel zur Reduktion von Selbstselektion.....	42
5.	Zustellung und Präsentation	44
5.1.	Zustellung des Fragebogens	44
5.2.	Präsentationsformen	44
5.3.	Situativer Charakter der Online-Befragung.....	46
6.	Erfassen der Antworten.....	48
6.1.	Datenspeicherung.....	48

6.2.	Kontrolle auf Eingabefehler und Konsistenz.....	49
6.2.1.	Fehlende und fehlerhafte Eingaben.....	49
6.2.2.	Konsistenzfehler.....	51
6.3.	Das Problem des Befragungsabbruchs.....	51
7.	Einsatzfelder für Online-Befragungen	53
7.1.	Generelle Befragungen der Online-Population.....	53
7.2.	Mitarbeiterbefragungen.....	54
7.3.	Nutzung und Akzeptanz von Web-Auftritten und Online-Dienstleistungen.....	54
7.4.	Fragebogenexperimente	54

1. Anatomie der Befragung

1.1. Definition

Als Befragung kann man alle empirischen Verfahren der Erhebung sozialer Realität verstehen, bei denen eine Einzelperson oder Gruppe auf mündlich oder schriftlich präsentierte Fragen (evtl. ergänzt durch weitere Stimuli) in mündlicher oder schriftlicher Form antwortet.

In der Regel wird die Befragung bei einer Stichprobe von Personen auf der Grundlage eines vorformulierten Fragebogens durchgeführt. Gängige Formen der Befragung sind das mündlich-persönliche Interview, die Telefon-, schriftlich-postalische und die Online-Befragung.

Befragungstypen lassen sich klassifizieren nach:

- dem Grad der Strukturierung des Fragebogens – bei stark strukturierten Befragungen folgt jedes Interview dem festgelegten Ablauf, bei schwach oder nicht strukturierten wird die Abfolge der Fragen durch den Gesprächsverlauf determiniert.
- dem Grad der Standardisierung – standardierte Befragungen folgen präzise den vorgegeben Formulierungen der Fragen, unstandardisierte orientieren sich stärker am Alltagsgespräch.
- dem Grad der Interaktivität – mündliche Befragungsformen implizieren soziale Interaktion zwischen einem Interviewer und einem Befragten, bei selbstauszufüllenden schriftlichen Befragungen fehlt die Interaktion.
- dem Distributionsmedium – mündlich im direkten Kontakt oder Telefon, schriftlich im direkten Kontakt, postalisch oder per computergestützter Kommunikation.
- dem Präsentationsmedium – verbal oder schriftlich – gedruckt oder über einen Computermonitor.

1.2. Befragung als Abfolge methodischer Schritte

Die praktische Durchführung lässt sich als eine Abfolge methodischer Schritte fassen, die in einem losen Abhängigkeitsverhältnis zueinander stehen. Die Um-

setzung der einzelnen Schritte ist abhängig vom Methodenverständnis, vor allem aber von den Rahmenbedingungen der Befragung und vom Erkenntnisinteresse.

- Operationalisierung – Umsetzung der Forschungsfragen oder Hypothesen in Fragen und Fragesequenzen – Frageformulierung und Fragebogenkonstruktion
- Ziehen der Stichprobe und Kontaktieren der Zielpersonen
- Zustellung und Präsentation des Fragebogens
- Erfassen der Antworten

Die Online-Befragung erhält ihre besondere Charakteristik aus der spezifischen Ausprägung der letzten drei Stufen:

- Der Fragebogen wird über ein Computernetzwerk (in der Regel das Internet) zugestellt.
- Der Fragebogen wird auf einem Bildschirm in schriftlicher Form präsentiert.
- Der Befragte beantwortet die Fragen eigenständig in schriftlicher Form.

Damit handelt es sich bei der Online-Befragung um eine Sonderform der schriftlichen Befragung: Man spricht von einem *Computer Aided Self Administered Questionnaire* bzw. einem *Computer Aided Self Administered Interview (CASI)*¹. Aus der Festlegung dieser Aspekte ergeben sich eine Reihe von Konsequenzen für die anderen Stufen der methodischen Umsetzung. Die folgenden Kapitel behandeln diese Besonderheiten anhand einer allgemeinen methodischen Diskussion der Befragungstechniken.

¹ Insofern die Eigenschaften der Online-Befragung betroffen sind, die sie mit anderen CASI-Formen teilt, wird die Abkürzung CASI synonym zu „Online-Befragung“ verwendet.

2. Frageformulierung und die Messung von Antwortverhalten

Am Beginn jeder empirischen Untersuchung steht die Formulierung von Forschungsfragen. Diese Forschungsfragen können gerichtet sein oder ungerichtet, von Theorien abgeleitet oder explorativer Natur, strukturiert oder unstrukturiert. Die empirische Studie dient, wenn die Forschungsfragen theoriegeleitete Hypothesen darstellen, dazu, die Gültigkeit der Annahmen anhand empirischer Beobachtungen zu überprüfen; in explorativen Studien sollen Beobachtungen über empirische Relative gemacht werden, die zur Entwicklung von gesetzesartigen Aussagen und Theorien über den beobachteten Realitätsausschnitt herangezogen werden können. Die Betrachtung unterschiedlicher Bereiche der Realität erfordert unterschiedliche Messinstrumente – oder eine Kombination aus mehreren Instrumenten.

Die Wahl des geeigneten Messinstruments sollte im Prinzip erst *nach* der Formulierung der Forschungsfragen erfolgen, nach einer sorgfältigen Abwägung, welches Instrument am besten geeignet ist². Die *methodische* Reflexion über die bestmögliche empirische Umsetzung setzt ein, nachdem die Wahl des Instruments getroffen wurde. Die erste und wichtigste Aufgabe besteht in der Operationalisierung, der Umsetzung der theoretisch begründeten Frage in die dem Instrument entsprechende Form.

Die Befragung stellt ein komplexes Messinstrument dar, mit dem sich Aussagen über das Denken, Fühlen und Handeln von Menschen treffen lassen. Diese drei Bereiche lassen sich nicht in gleichem Maße mit Hilfe von Fragen messen – so sind Aussagen über vergangenes Handeln von der Erinnerungsfähigkeit der befragten Person abhängig, Aussagen über Gefühle von ihrer Artikulationsfähigkeit. Die Operationalisierung muss dieser unterschiedlich ausgeprägten Zugänglichkeit menschlicher Erfahrungs- und Handlungswelten Rechnung tragen.

Die Formulierung von Forschungsfragen erfolgt auf der Ebene theoretischer Konstrukte – kognitive und Handlungs-Muster, Meinungen und Einstellungen. Bei der methodischen Umsetzung ist darauf zu achten, dass diese Konstrukte

² In der Praxis ist dies häufig nicht möglich, da externe Umstände die Wahl des Instruments mit bestimmen.

durch die Messinstrumente – Fragen und Fragesequenzen – valide abgebildet werden.

Insgesamt umfasst die Operationalisierung folgende Aspekte:

- Formulierung von Fragen;
- Festlegung der Antwortmöglichkeiten – und damit der Messvorschriften für die Äußerungen des Befragten;
- Konstruktion von Skalen aus Fragesequenzen;
- Zusammenstellung der Fragen und Fragesequenzen zu einem Fragebogen.

2.1. Frageformulierung

2.1.1. Anforderungen an die Frage

Die Frage, die dem Befragten vorgelegt (oder –gelesen) wird, soll ihn veranlassen, über einen bestimmten Bereich seines Denkens, Fühlens, Handelns Auskunft zu geben. Dazu ist es notwendig, dass

- der Befragte die Frage so versteht, dass er die Intention des Forschers nachvollziehen kann, die Frage also im Sinne des Forschungsvorhabens „richtig“ versteht;
- er *in der Lage* ist, die Frage zu beantworten;
- er *bereit* ist, die Frage – wahrheitsgemäß – zu beantworten.

Da in der Regel nicht nur eine Person befragt wird, ist es außerdem notwendig, dass die beschriebenen kognitiven und motivationalen Voraussetzungen für alle Befragten möglichst in gleichem Maße gelten. Die erste, individuelle Dimension dieser Anforderungen entspricht der Forderung nach der Validität eines Messinstrumentes – es soll das messen, was es im Kontext der zugrunde liegenden Theorie messen soll, die zweite, kollektive, Dimension betrifft die Reliabilität – die Fähigkeit des Instruments, das zu erfassende Phänomen unabhängig von störenden Randbedingungen immer in gleicher Art und Weise zu messen. Für die Frageformulierung lassen sich daraus zunächst *kognitive* Anforderungen ableiten:

Eine Frage soll von jedem Befragten in der gleichen – vom Forscher intendierten – Art und Weise möglichst sofort verstanden werden. Wenn wir davon ausgehen, dass in der befragten Population die kognitiven Fähigkeiten als Grundlage des Frageverstehens variieren, muss die Frage also möglichst einfach zu verstehen sein. Konkret bedeutet dies:

- Verwendung einfacher Wörter;
- Vermeidung von Fremdwörtern;
- einfacher Satzbau, kurze Sätze, möglichst keine Nebensätze oder missverständliche temporale oder konjunktive Konstruktionen.

Natürlich muss man bei der Umsetzung dieser Regeln die zu untersuchende Population berücksichtigen. Man kann – und muss – davon ausgehen, dass die Fähigkeit zum Verständnis komplexer Fragen mit der formalen Bildung korreliert – so wird man z.B. bei einer Befragung von Journalisten die Komplexitätsschwelle höher ansetzen können, als bei einer Bevölkerungsumfrage. Man wird im ersten Fall sogar auf zu einfache Konstruktionen verzichten – der Befragte könnte sich „für dumm verkauft“ fühlen.

Lässt man die Anforderung adäquater Einfachheit außer acht, so wird die Reliabilität des Instruments darunter leiden: Nicht alle Befragten reagieren in gleicher Art und Weise, wenn ihr Verständnis der Frage variiert. Eine zu einfache Formulierung kann dagegen die Validität beeinflussen – wenn die Simplifizierung die beabsichtigte Fragestellung nicht mehr adäquat widerspiegelt.

Bsp:

Frage A: Glauben Sie, dass der öffentlich-rechtliche Rundfunk das Prinzip des Binnenpluralismus ausreichend berücksichtigt?

Frage B: Glauben Sie, dass im öffentlich-rechtlichen Rundfunk alle ihre Meinung äußern können?

Frage A ist für viele Befragte nicht zu beantworten, weil sie mit dem Begriff „Binnenpluralismus“ nichts anfangen können; Frage B bildet das Konzept „Binnenpluralismus“ nicht ausreichend ab.

In vielen Fällen ist es nicht möglich, eine schwierig zu verstehende Frage durch eine einfache zu ersetzen, statt dessen muss der komplexe Sachverhalt in meh-

rere einzelne Aspekte – und damit Fragen – aufgeteilt werden: in eine *Skala*, deren Komponenten zusammengenommen den Sachverhalt messen können.

Der zweite Bereich von Anforderungen ist *motivationaler* Natur: Der Befragte muss nicht nur fähig, sondern auch willens sein, die Frage zu beantworten. Dazu müssen *motivierende* Faktoren in den Vordergrund gerückt, *hemmende* möglichst unterdrückt werden:

- der Befragte soll sich angesprochen fühlen, zu der Frage durch die Wiedergabe seiner Meinung oder seines Wissens Stellung zu nehmen.
- der Befragte darf sich durch die Frage nicht persönlich angegriffen oder in seiner Privatsphäre beeinträchtigt fühlen.

Eine weitere Anforderung betrifft die *rhetorische* Qualität: Die Frageformulierung darf den Befragten nicht beeinflussen, die Frage in einer bestimmten Art und Weise zu deuten und zu beantworten³.

- Suggestivfragen, die eine bestimmte Antwort nahelegen, sind zu vermeiden.

Bei der Formulierung von Fragen muss beachtet werden, dass der Befragte die Fragen innerhalb einer spezifischen Kommunikationssituation wahrnimmt. Diese Wahrnehmung wird unter anderem dadurch beeinflusst, als Repräsentant welcher Population sich der Befragte angesprochen fühlt⁴. Weitere Faktoren sind die Präsenz eines Interviewers, die Präsentation in Gesprächs- oder in schriftlicher Form, die Form der Aktivität, die der Befragte zeigen muss, um die Fragen zu beantworten. Es ist notwendig, die Frageformulierung auf diese Situation hin anzupassen. Darüber hinaus gibt es eine Reihe weiterer Faktoren, die sich vom Forscher nicht kontrollieren lassen – z.B. der aktuelle Gemütszustand des Befragten und seine Einstellungen gegenüber Umfragen.

³ Oder, anders formuliert, sie darf dem Befragten nicht die Angemessenheit einer bestimmten, durch den Forscher vorgegebene Interpretation sozialer Realität vorgeben, innerhalb derer bestimmte Antworten positiv oder negativ evaluiert wären.

⁴ In einer Umfrage unter Fachleuten wird ein Naturwissenschaftler eine laienhaft formulierte Frage als unangemessen betrachten, weil er sich als Experte angesprochen fühlt. Bei einer allgemeinen Bevölkerungsumfrage wird er über die identische Formulierung eher hinwegsehen.

2.1.2. Frageformulierung: CASI

Die oben formulierten Grundregeln gelten für schriftliche ebenso wie für mündliche Befragungsformen. Gegenüber der mündlichen bietet die schriftliche Befragung den Vorteil, dass der Befragte die Möglichkeit hat, die Frage mehrfach zu lesen und über ihre Bedeutung nachzudenken. Dem steht der Nachteil der fehlenden Interaktion gegenüber: kein Gesprächspartner steht für Nachfragen zur Verfügung. Bei einer schriftlichen Befragung kommt daher der begrifflichen Klarheit eine besondere Bedeutung zu. Insbesondere bei Online-Befragungen sollte man darauf achten, die Geduld des Befragten nicht mit zu langen oder komplexen Fragen zu strapazieren⁵. Erläuterungen sind auf das Notwendige zu reduzieren. Durch die Möglichkeit, Zusatzinformationen (z.B. Begriffserklärungen) anzubieten, die der Befragte bei Bedarf abrufen kann, wird das Manko fehlender Interaktion – zumindest bzgl. des möglichen Ausgleichs von Informationsdefiziten – teilweise ausgeglichen.

2.2. Antworten und ihre Messung

2.2.1. Messung

Um soziale wie physikalische Phänomene zu erfassen, nimmt der Forscher eine Messung vor. Bei der Befragung wird die Reaktion einer Person auf eine Frage gemessen⁶. Diese Reaktion stellt eine Handlung dar – in einer standardisierten Befragung lässt sich diese Handlung nicht in ihrer Gänze, es lassen sich nur bestimmte Aspekte dieser Handlung messen.⁷ Die quantitative Sozialforschung lehnt sich bei den Anforderungen an Messungen an die Naturwissenschaft an, mit der Forderung nach:

- Isomorphie: Strukturelle Ähnlichkeit des Messinstruments mit dem Gegenstand der Messung.

⁵ Die Zeit, die ein Befragter beim Ausfüllen eines Online-Fragebogen verbringt, kostet ihn Geld – zumindest dann, wenn er eine private Internetverbindung benützt, die im Zeittakt abgerechnet wird. Ausschweifende Frageformulierungen erhöhen also bei einem Teil der Population die Gefahr eines Abbruchs der Befragung und damit das Entstehen systematischer Verzerrungen.

⁶ Die Gegenüberstellung Reiz – Reaktion ist lediglich als Mittel der Strukturierung des Ablaufs eines Interviews zu verstehen, nicht im Sinne einer inhaltlichen Einordnung als Stimulus-Response-Prozess.

⁷ Im Kern der Diskussion um qualitative vs. quantitative Umfrageforschung steht die Frage, ob eine systematische Reduktion dieser Sprach-Handlung auf bestimmte, vom Forscher vorgegebene Dimensionen der Messung legitim ist.

- Validität: Die Messung soll das Phänomen erfassen, das vom Forscher theoretisch spezifiziert wurde.
- Reliabilität: Die Messung soll unter verschiedenen Rahmenbedingungen immer die gleichen, möglichst präzisen Ergebnisse liefern. Reliabilität setzt sich zusammen aus:
 - Stabilität
 - Intersubjektivität
 - Genauigkeit

Im Rahmen einer Befragung kann man dann von einer Messung sprechen, wenn die Antwort des Befragten nicht einfach originalgetreu aufgezeichnet wird (offene Fragen) sondern in irgendeiner Form klassifiziert. Messung bedeutet immer eine Umsetzung und Reduktion der Wahrnehmung des Untersuchungsgegenstands. In der Regel wird in einer Befragung die Klassifikation vorgegeben, d.h. dem Befragten werden von vornherein eingeschränkte Auswahlmöglichkeiten präsentiert. Die Klassifikation kann aber auch – im Falle eines persönlichen Interviews – durch den Interviewer erfolgen, oder im Nachhinein auf der Basis der protokollierten Antwort.

2.2.2. Das Universum möglicher Antworten: Skalentypen

Eine grundlegende Unterscheidung von Antworttypen im Rahmen einer Befragung ist die in *geschlossene* und *offene* Fragen. Offene Fragen ohne Antwortvorgaben eignen sich, um die uneingeschränkte Reaktion des Befragten zu erfassen. Im Gegensatz zu einem Alltagsgespräch oder einem journalistischen Interview wird die sozialwissenschaftliche Befragung das Universum möglicher Antworten meist einschränken. Diese Einschränkung ist mit einem Informationsverlust verbunden – der Forscher postuliert, dass hier nützliche von unnützer Information getrennt wird. Diese Trennung ist essentiell, um die Eindimensionalität der Messung zu gewährleisten. Die Qualität einer Frage ist davon abhängig, inwieweit der Anspruch auf den Erhalt nützlicher Information eingehalten werden kann.

Bsp.

Die Frage „Wie beurteilen Sie den Fernsehsender *ProSieben*?“ kann, wenn sie als offene Frage gestellt wird, eine Vielzahl möglicher Antworten aus-

lösen, die nicht nur die unterschiedliche Bewertung, sondern auch verschiedene Erfahrungsweisen und verbale Konventionen abbilden. Ein Universum möglicher Antworten, das „toll“, „anspruchlos“, „so was schau ich mir nicht an“ und mehr umfassen kann, macht den Vergleich der Antworten vieler Respondenten außerordentlich kompliziert, denn in ihnen sind viele, mehrere Dimensionen umfassende, Informationen enthalten. Die Verwendung einer Skala „sehr gut“ – „gut“ – „mittelmäßig“ – „schlecht“ – „sehr schlecht“ zwingt die Antworten in eine Dimension, die einen einfachen Vergleich ermöglicht. An dem Beispiel lässt sich auch ablesen, dass die Einschränkung auf eine Dimension nicht immer sinnvoll ist – zum Beispiel, wenn man eben diese Mehrdimensionalität der Wahrnehmung und Bewertung zum Untersuchungsgegenstand hat.

Je nach dem zur Untersuchung stehenden Gegenstandsbereich wird man unterschiedliche Strategien zur Einschränkung des Antwort-Universums verwenden.

Die einfachste Form ist die Umsetzung in Form einer „ja/nein“-Antwort: eine *dichotome* Skala.

Bsp.: Besitzen Sie ein Fernsehgerät? (ja/nein)

Zur Erhebung von Differenzen, die über eine binäre Unterscheidung hinausgehen, verwendet man eine *nominale* Skala, die Unterscheidung zwischen mehreren Objekten ermöglicht.

Bsp.: Auf welche Weise empfangen Sie Fernsehen? (Antenne/Kabel/Satellit)

Die Einordnung von Zuständen bezüglich einer Bewertungsdimension kann im einfachsten Fall in Form einer *ordinalen* oder Rangskala (Ranking) geschehen. Ein Sonderfall der Rangskala ist der *Paarvergleich*, bei dem die Rangfolge aus dem direkten Vergleich von je zwei Objekten rekonstruiert wird.

Bsp.: Ordnen Sie folgende Sender danach, wie häufig Sie sie einschalten. (ARD/ZDF/SAT.1/RTL)

Eine Einschätzung hinsichtlich eines abstrakten Maßstabs (Rating) auf einer *Intervall*-Skala liefert die informationsreichsten Daten, denn die hier verwendeten numerischen Werte drücken nicht nur Unterschiedlichkeit aus, wie bei der Nominalskala, oder eine Reihenfolge, wie bei der Ordinalskala: Numerische

Unterschiede bilden hier empirische Unterschiede proportional ab. Allerdings ist es häufig zweifelhaft, ob im Rahmen einer Befragung die Anforderungen an eine Messung auf Intervallniveau im strengen Sinn zu erfüllen sind, da man nicht ohne weiteres davon ausgehen kann, dass alle Befragten die gleichen Maßstäbe anlegen und gleiche psychologische Differenzen in gleiche numerische Messwerte umsetzen⁸. Aufgrund der eingeschränkten statistischen Auswertungsmöglichkeiten bei ordinalskalierten Daten und der meist geringen Gefahr substantieller Fehlschlüsse aufgrund der Überschätzung des Datenniveaus, ist es durchaus üblich, Daten, die streng genommen nur ordinales Skalenniveau beanspruchen können, als intervallskaliert zu behandeln, wenn sie akzeptable Verteilungsparameter aufweisen⁹.

Bsp.: Beurteilen Sie die Nachrichtensendung *heute* auf einer Skala von 1 (sehr gut) bis 5 (sehr schlecht).

Insbesondere in der Einstellungsmessung spielen Rating-Skalen eine außerordentlich wichtige Rolle. Auch wenn die Umsetzung des Verfahrens grundsätzlich einfach erscheint, gibt es doch eine Reihe von Aspekten, die Beachtung verdienen:

- *Anzahl der Skalenpunkte*: Je mehr Unterteilungen eine Skala besitzt, desto detaillierter ist die Messung. Da Messgenauigkeit ein Kriterium für die Güte der Messung ist, ließe sich daraus folgern: je differenzierter, desto besser. Dies gilt aber nicht uneingeschränkt: Selbst bei physikalischen Messungen geht man davon aus, dass es sinnlos ist, eine Messung in einer Einheit auszuweisen, die kleiner ist als der Messfehler (z.B. Städteentfernungen in cm). Bei psychometrischen Verfahren spielt der Messfehler eine noch größere Rolle. Von zentraler Bedeutung ist hier die *Fähigkeit* des Rezipienten zur Einschätzung. So ist z.B. ein Weinexperte wie Robert Parker in der Lage, die Qualität eines Weines auf einer Skala von 50 bis 100 Punkten einzuschätzen¹⁰. Das differenzierte Urteil mag bei Fachleuten umstritten sein, wird aber in der Regel als begrün-

⁸ So ist z.B. bei einer Benotung die psychologische Differenz zwischen einer 4 und einer 5 größer als der zwischen einer 3 und einer 4 – im ersten Fall wird zwischen "bestanden" und "nicht bestanden" unterschieden, im zweiten liegt nur eine qualitative Abstufung der Bewertung vor.

⁹ Eine Vielzahl statistischer Verfahren beruhen direkt oder indirekt auf dem Begriff der Varianz und setzen approximativ normalverteilte Daten voraus.

¹⁰ vgl. <http://www.erobertparker.com/info/legend.asp>

det akzeptiert, eine vergleichbare Einschätzung eines Laien in diesem Gebiet wäre dagegen aufgrund dessen mangelnder Erfahrung und Differenzierungsfähigkeit völlig inakzeptabel. Würde man vergleichbare Skalen in Befragungen anwenden, so wären geringe Differenzen zwischen einzelnen Personen nicht das Ergebnis der unterschiedlichen zugrundeliegenden „wahren“ Ausprägungen, sondern Fehlervarianz. Ein weiterer Nachteil differenzierter Skalen liegt in der höheren Belastung des Befragten, der sich mit der Aufgabe konfrontiert sieht, übermäßig differenzierte Einschätzungen vorzunehmen.

Am anderen Ende des Kontinuums liegen Skalen, die zu wenig Ausprägungen aufweisen – und damit die Varianz des zugrundeliegenden Merkmals nicht mehr adäquat abbilden. Mit simulierten Daten lässt sich nachweisen, dass Skalen mit vier Ausprägungen oder mehr die Varianz einer zugrundeliegenden kontinuierlichen Skala noch akzeptabel wiedergeben. Die häufig verwendeten Skalenumfänge von 5 – 7 sind ein sowohl statistisch als auch psychologisch guter Kompromiss.

- *Gerade vs. ungerade Anzahl von Skalenpunkten:* Die Existenz eines Mittelpunkts einer Skala repräsentiert die Existenz einer neutralen Position. Die Tendenz vieler Befragte, Mittelpositionen anzukreuzen, wird dadurch noch verstärkt. Insbesondere, wenn die Endpunkte gegensätzliche Positionen ausdrücken, liegt bei unsicheren Personen die Flucht in die mittlere Position nahe. Häufig wird dieser Skalenpunkt auch als Synonym für „weiß nicht“, oder „keine Meinung“ missbraucht.

Andererseits gibt es Fragen, bei denen eine valide neutrale Position existiert. Das Fehlen einer mittleren Position würde vom Befragten dann mit einer zufälligen Auswahl einer der mittleren Kategorien quittiert werden.

2.2.3. Antworterhebung: CASI

Generell lassen sich in Online-Befragungen sämtliche gängigen Formen von Antwortvorgaben verwenden. Sie teilen mit anderen Typen schriftlicher Befragungen Stärken und Schwächen. Im Gegensatz zu anderen Befragungssituationen, bei denen kein Interviewer zugegen ist, erlaubt die Online-Befragung eine

gewisse Kontrolle über das Antwortverhalten, die dazu beiträgt, fehlerhafte Angaben zu vermeiden.

- **Offene Fragen** erfordern vom Befragten, eine eigene Antwort einzutragen.
 - Stärken bei CASI: Die Antwort kann nicht durch Probleme der akustischen Verständlichkeit verfälscht werden, der Befragte kann Umfang und Geschwindigkeit der Antwort selbst wählen. Gegenüber Paper&Pencil-Befragungen hat die Online-Befragung den Vorteil, dass bei der Auswertung der Antworten keine Handschrift entziffert, kein Text neu erfasst werden muss. Der Antworttext steht für die weitere Analyse (z.B. durch computergestützte Inhaltsanalyseverfahren) sofort zur Verfügung.
 - Schwächen: Der Aufwand für den Befragten ist bei einer schriftlichen offen formulierten Antwort höher als bei der mündlichen. Dies kann seine Bereitschaft zu einer vollständigen Auskunft einschränken. Beim Befragten wird Schreibfähigkeit vorausgesetzt (dies schließt Populationen aus, die aufgrund des Alters oder einer Behinderung nicht schreiben können). Die computergestützte Beantwortung setzt natürlich ein Mindestmaß an „Computer Literacy“ voraus, deren unterschiedliche Ausprägung die Antwortbereitschaft beeinflussen kann – die souveräne Beherrschung der Tastatur macht eine ausführlichere Antwort einfacher¹¹. Gegenüber Paper&Pencil schränkt der Computer nicht-textuelle Antwortmöglichkeiten ein, z.B. das Aufzeichnen einer Situation oder eines Gegenstands – dies ist nur mit erheblichem Programmieraufwand und unter erheblichen Voraussetzungen bei der technischen Ausstattung möglich. Metalinguistische Aspekte (Dialekte, Jargons) werden durch die schriftliche Antwort nivelliert, gegenüber der handschriftlichen Eingabe lässt diejenige über die Tastatur auch keine Rückschlüsse auf den Befragten über seinen Schriftstil zu.
 - Technische Probleme: Eine Eingabebeschränkung für die Länge der Textfelder wird meist sinnvoll sein, muss dem Befragten aber auch mitgeteilt werden: Ist das Eingabefeld länger als das Feld in der Datenbank, in der die Informationen erfasst werden, geht unter Umständen wertvolle Information verloren.

¹¹ Die Anforderung des Zugangs zu einem Computer engt natürlich bei der Befragung einer Bevölkerungstichprobe die Population von vornherein stark ein. Aber auch bei der Befragung in einer Laborumgebung muss man berücksichtigen, ob die Befragungsteilnehmer in vergleichbarem Ausmaß über Fähigkeiten im Umgang mit dem Computer verfügen.

- **Nominale Skalen** (inkl. dichotome) erfordern vom Befragten die Auswahl einer oder mehrerer zutreffender Antworten.
 - Stärken: Ein Problemfeld des Selbst-Ausfüllens in der schriftlichen Befragung besteht in der mangelnden Kontrolle des Verhaltens des Befragten – so kann dieser aus Versehen mehrere Antworten eintragen, wo keine vorgesehen sind. In der computergestützten Befragung lässt sich aber die Anzahl der ausgewählten Felder genau kontrollieren: Im einfachsten Fall (bei genau einer Antwortmöglichkeit), dadurch, dass die Auswahl einer Alternative die zuvor gewählte aufhebt. Bei Mehrfachantworten lässt sich durch das Überprüfen der Anzahl der gegebenen Antworten kontrollieren, ob der Befragte in dem vorgegebenen Bereich bleibt. Korrekturen der Antwort durch den Befragten sind möglich, ohne dass Missverständnisse bei der Auswertung entstehen können. Durch den programmgestützten Ablauf lassen sich komplexe Filterführungen auch auf der Basis von Mehrfachantworten realisieren – bei papiergestützten Verfahren würde dies den Befragten überfordern.
- **Rangskalen:** Der Befragte ordnet jedem Objekt genau einen Rangplatz zu.
 - Stärken: Bei Pencil&Paper-Befragungen sind bei einer Rangbildung Fehler möglich, bzw. kann der Befragte selbständig entscheiden, ob er sich einer Entscheidung entziehen möchte, indem er einen Rangplatz doppelt vergibt. Während in der mündlichen Befragung der Interviewer als Kontrollinstanz eingreifen kann, ist in der computergestützten Befragung die korrekte Vergabe per Programm kontrollierbar. Bei Paarvergleichen ist es darüber hinaus möglich, inkonsistente Angaben¹² bereits bei der Angabe herauszufiltern.
- **Ratingskalen:** Der Befragte vergibt auf einer Skala für ein Objekt Werte bezüglich eines Beurteilungskriteriums.
 - Stärken: Die versehentliche Vergabe mehrerer Werte für ein Objekt kann ausgeschlossen werden, ebenso unklare Angaben.
 - Technische Probleme: Die Umsetzung von graphischen Alternativen zu Skalen, auf denen der Befragte eine genaue Position einzeichnen kann (z.B. in Form eines Thermometers) ist schwierig und aufgrund möglicher Darstellungsprobleme nur in einer Laborsituation zu empfehlen, in der man die Kontrolle über die Präsentation des Fragebogens besitzt.

¹² In jedem Paarvergleich sind zirkuläre Angaben der Art $a > b$, $b > c$, $c > a$ denkbar, obwohl sie in den meisten Fällen von der Theorie her als nicht gültig angesehen werden dürften.

2.3. Itemlisten und Skalierungsverfahren

2.3.1. Itemlisten

In einer Befragung tritt häufig der Fall auf, dass zu einem Objekt eine Reihe von Attributen abgefragt wird. In diesem Fall ist es wenig sinnvoll, jede Objekt-Attribut-Kombination in einer eigenen Frage zu behandeln. Statt dessen wird in der Frageformulierung darauf hingewiesen, dass zu diesem Objekt verschiedene Eigenschaften abgefragt werden, die jeweils auf der gleichen Skala zu bewerten sind.

Bsp:

Hier sind einige Eigenschaften, die man der "Tagesschau" zuordnen kann. Bitte geben Sie jeweils an, wie sehr Sie damit übereinstimmen, dass die Eigenschaft auf die "Tagesschau" zutrifft.

Die Tagesschau ist ...

langweilig () ...

aufregend

informativ

Das gleiche Prinzip ist anzuwenden, wenn es um ein Attribut verschiedener Objekte geht:

Bsp.

Wie oft informieren Sie sich aus folgenden Quellen über das aktuelle Tagesgeschehen?

Fernsehen () ...

Tageszeitung

Internet

2.3.2. Skalierung

Komplexe psychische Sachverhalte lassen sich nicht durch einzelne Fragen messen. Die *klassische Testtheorie* geht davon aus, dass sich derartige Sachverhalte als latente Variablen beschreiben lassen, die nicht direkt, aber indirekt durch mehrere manifeste Indikatoren sind. Der wahre Wert einer latenten Variable

ergibt sich durch eine gewichtete Linearkombination dieser Indikatoren plus einem residualen Fehler.

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n + e.$$

Eine Skala, die diesem Schema folgt, besteht aus einzelnen Indikatoren, die untereinander relativ hoch korreliert ist. Ein wesentliches Kriterium ist die *Ein-dimensionalität*, d.h. die Skala soll ein einziges Phänomen messen, das in unterschiedlicher Ausprägung vorliegen kann.

Skalierungsverfahren

Die in der Theorie verwendeten Hypothesen benutzen Konstrukte, die in der Testtheorie durch latente Variablen repräsentiert werden. Ein zentrales Problem der Operationalisierung besteht nun darin, die richtigen Indikatoren zu finden, um eine valide und reliable Messung zu ermöglichen.

Wenn es um Konstrukte geht, die bereits Gegenstand sozialwissenschaftlicher Forschung waren, sollte man auf vorhandene, überprüfte Skalen zurückgreifen, die man in Handbüchern wie dem Zuma-Skalenhandbuch findet. Bei Neuentwicklungen ist es notwendig, Verfahren einzusetzen, die eine möglichst ökonomische Konstruktion von Skalen ermöglichen. Die Reliabilität dieser Skalen lässt sich dann mit Hilfe von Verfahren wie der Item- und der Faktorenanalyse ermitteln.

Verfahren nach Thurstone

Die Thurstone-Skala ist das Ergebnis eines zweistufigen Prozesses. In der ersten Stufe wird eine große Zahl von Statements ausgewählt, von denen der Forscher annimmt, dass sie das zu untersuchende Phänomen repräsentieren. Diese Liste wird einer Gruppe von Experten vorgelegt, die auf einer mehrstufigen Skala einordnen, in welchem Ausmaß das jeweilige Statement das zugrundeliegende Konzept repräsentiert. Die Statements, bei denen die größte Übereinstimmung herrscht, werden für die zweite Stufe – die eigentliche empirische Untersuchung – ausgewählt. Die Probanden müssen nur noch Zustimmung oder Ablehnung zu den Statements äußern, bei der Konstruktion der Skala aus den einzelnen Items werden sie mit dem zuvor bestimmten Skalenwert gewichtet.

Durch die Validierung, die die Items durch die diversen Expertenmeinungen erhalten, besitzt die Skala durchaus theoretische Vorzüge. Sie wird aber nur

noch selten angewendet – vor allem aufgrund des Aufwands, der in der ersten Stufe betrieben werden muss. Durch die Möglichkeit, das Internet zur Kontaktaufnahme und Datensammlung in der Konstruktionsphase zu benutzen, lässt sich dieser Aufwand aber auf ein akzeptables Maß reduzieren.

Verfahren nach Likert

Das von Rensis Likert entwickelte Verfahren geht ebenfalls mehrstufig vor. In einem Pretest wird eine große Anzahl von Items, die vom Forscher formuliert wurden, von Befragten auf einer mehrstufigen Skala beantwortet. Die Items, die zu einem Konstrukt gehören, werden einer statistischen Analyse unterzogen: Im einfachsten Fall wird die Item-Total-Korrelation für jedes Item berechnet – also die Korrelation zwischen dem einzelnen Item und der Summe der Items. Nur die stark korrelierenden Items werden für die Hauptuntersuchung beibehalten.

Beim Einsatz der Likert-Skala haben sich Vereinfachungen eingebürgert. Auf die Voruntersuchung wird meist zugunsten einer heuristischen Vorauswahl verzichtet, schwach korrelierende Items werden nachträglich ausgeschlossen¹³.

Häufig findet im Kontext der Skalenkonstruktion nach Likert die Faktorenanalyse Anwendung, um zu überprüfen, ob eine Vielzahl von Items mehrere unabhängige Dimensionen repräsentieren.

Verfahren nach Guttman

Im Gegensatz zu den Skalierungsverfahren nach Likert und Thurstone, bei denen die Items zu einer Skala addiert werden geht das Verfahren nach Guttman von der Überlegung aus, dass die Zustimmung zu unterschiedlichen Items verschiedene Positionen auf einer zugrundeliegenden Skala markiert. So lassen sich die Statements „Ich gehe regelmäßig wählen“, „Ich informiere mich täglich über politische Ereignisse“ und „Ich bin aktives Mitglied in einer Partei“ als zunehmend hohe Ausprägungen auf einer Skala des politischen Involvements betrachten: Der Skalenwert ergibt sich aus der Zustimmung zu dem Item mit der höchsten Wertigkeit

¹³ Eine weitere Degeneration ist in der Begriffsverwendung „Likert-Skala“ zu beobachten. Die mehrstufigen Einzelskalen der Items werden häufig als „Likert Type Scale“ bezeichnet, was im deutschen Sprachgebrauch dann zu Likert-Skala verkürzt wurde.

Semantisches Differential

Das Semantische Differential wurde von Osgood, Suci und Tannenbaum entwickelt, um ein generelles Instrument für die Einschätzung von Objekten zu erhalten. Die Autoren gehen davon aus, dass jede Einschätzung auf drei grundlegenden Dimensionen beruht: Stärke, Aktivität und Bewertung. Bei der Konstruktion des SD werden Gegensatzpaare verwendet, die diese drei Dimensionen repräsentieren. Mit Hilfe der Faktorenanalyse wird überprüft, ob die Zuordnung der Items zu den Dimensionen statistisch haltbar ist.

Polaritätenprofil

Die Beschreibung von Objekten oder Sachverhalten durch Gegensatzpaare wird inzwischen auch häufig angewendet, ohne dass ein expliziter Bezug auf semantische Differenziale vorhanden ist. Da die Einordnung auf vergleichbaren Skalen wie im Verfahren nach Likert erfolgt, lassen sich hier entsprechende statistische Methoden einsetzen, um zu summierten Skalenwerten zu gelangen. Meistens ist aber weniger der Summenwert interessant, als das Bewertungsprofil verschiedener Objekte.

3. Fragebogenkonstruktion

Einzelfragen und Skalen stellen lediglich einen Teil eines komplexen Messinstruments dar. Die erfolgreiche Durchführung einer Befragung hängt in gleichem Maße von der richtigen Anordnung der Fragen in längeren Sequenzen wie von ihrer Formulierung ab. Fragebögen müssen einer Dramaturgie folgen, die von dem Befragten als sinnvoll akzeptiert wird. Durch die Struktur des Fragebogens soll außerdem sichergestellt werden, dass Ausstrahlungseffekte und sequentielle Fehler minimiert werden.

Bei der Betrachtung Struktur ist zu unterscheiden zwischen der Makro- und der Mikrostruktur. Unter der Makrostruktur ist die Anordnung längerer zusammenhängender Fragesequenzen zu verstehen. Die Mikrostruktur betrifft die Zusammenstellung von Fragen und Itemlisten/Skalen zu Fragesequenzen.

3.1. Makrostruktur

3.1.1. Einleitung

Jeder Fragebogen – insbesondere wenn er vom Befragten selbst ausgefüllt wird – sollte durch Einleitung und Verabschiedung eingerahmt sein.

Die Einleitung hat zwei Hauptfunktionen: Zum einen soll sie den Befragten motivieren und Vertrauen schaffen. In einer persönlichen Befragung kommt diese Aufgabe dem Interviewer zu, der den Vorteil hat, mit dem Befragten bereits in eine soziale Situation eingetreten zu sein, die dieser nicht ohne weiteres abbrechen wird. Die persuasive Funktion der Einleitung in einer schriftlichen Befragung muss dagegen allein durch die rhetorische Qualität der Einleitung erfüllt werden – denn die Zielperson kann jederzeit die Rezeption abbrechen und als potentieller Befragter verloren gehen.

Die Einleitung sollte folgende Elemente enthalten, die die wichtigsten Fragen, die bei der Zielperson, die mit der Befragung konfrontiert wird, auftauchen werden, zu Beginn beantwortet:

- „Warum?“ Eine kurze Schilderung des Forschungsinteresses, das hinter der Befragung steht. Je konkreter diese Beschreibung ist, desto eher wird sie auch in der Lage sein, beim Befragten Interesse zu erwecken. Allerdings kann dies auch dazu führen, dass die Rolle spezifischer Interessen die Selbstselektion verstärkt. Eine zu ausführliche Beschreibung

kostet außerdem viel Zeit, und schließlich können zu präzise Vorstellungen des Befragten über das Ziel der Befragung Einfluss auf sein Antwortverhalten ausüben.

- „Wer?“ Bei einem rein universitären Forschungsvorhaben ist es unproblematisch, dem Befragten mitzuteilen, wer die Untersuchung durchführt. Kommerzielle Interessen komplizieren die Angelegenheit: Hier könnte die genaue Nennung des Auftraggebers die Bereitschaft zur Teilnahme und das Antwortverhalten beeinflussen. Besonders problematisch ist die Lage bei universitärer Auftragsforschung: Das Verschweigen der kommerziellen Interessen hinter einer Umfrage ist nicht nur aus wissenschaftsethischer Sicht absolut inakzeptabel, es schadet längerfristig auch dem Renommee universitärer Forschung.
- „Wie lange?“ Der Befragte sollte eine realistische Vorstellung davon erhalten, wie lange die Befragung dauert. Da es keine Möglichkeit gibt, den Befragten über einen bestimmten Zeitpunkt hinaus zur Teilnahme zu bewegen, ist hier eine Unterschätzung der Befragungsdauer kontraproduktiv und kann schnell zum Abbruch der Befragung führen.
- Vertraulichkeit: Die Zusicherung der Vertraulichkeit ist in der Umfrageforschung selbstverständlich. Dem Befragten muss glaubhaft gemacht werden, dass kein Versuch erfolgt, seine Antworten mit ihm als Person in Beziehung zu bringen und dass keine de-anonymisierten Daten Dritten zur Verfügung gestellt werden. Im Kontext der Online-Befragung sind einige kritische Aspekte erwähnenswert:
 - Die Form der Kontaktaufnahme: Bei einer passiven Stichprobe muss sich der Befragte um seine „Anonymität“ zunächst keine Gedanken machen – selbst die Ermittlung der verwendeten IP-Adresse reicht im Normalfall nicht zur Identifikation der teilnehmenden Person. Erfolgt die Kontaktaufnahme aktiv über E-Mail, muss dem Befragten deutlich werden, dass der Link, der ihn von der Mail zum Fragebogen führt, keine Informationen enthält, die zu seiner Identifikation führen. Technisch wäre dies möglich: So könnte in eine HTML-formatierte Mail etwa ein Link der Form
`Zum Fragebogen` einge-

fügt und die Fragebogen-Seite in einem Frameset versteckt werden um die Identifikationsnummer zu kaschieren¹⁴.

- Ein weiteres Problem stellen Cookies dar: In einem komplexen Fragebogen ist die Verwendung von Cookies nur schwer zu umgehen. Diese Cookies können aber auch Informationen speichern, die zu einem späteren Zeitpunkt zu kommerziellen Zwecken missbraucht werden könnten¹⁵. Nutzer, die sich beim Setzen von Cookies vom Browser warnen lassen, könnten misstrauisch werden, wenn der Zweck der Cookies nicht erläutert wird.
- Incentives: Die Verwendung von Incentives materiellen Charakters bedeutet bei einer Online-Befragung fast zwangsläufig den Bruch der Anonymität. Hier muss darauf geachtet werden, dass das Verfahren zur Inanspruchnahme der Incentives vom Fragebogen streng getrennt ist – aber auch dann ist das Zusammenführen von persönlichen Daten im Rahmen der Incentive-Vergabe und den Fragebogen-Daten technisch möglich. Eine Vertrauensgrundlage zwischen Forscher und Befragtem ist hier von großer Bedeutung. „Immaterielle“ Incentives bilden hier eine interessante Alternative, z.B. die Zugriffsmöglichkeit auf attraktive Web-Inhalte nach dem Ausfüllen des Fragebogens (vgl. 4.4.3).

Eine wichtige Funktion der Einleitung besteht in der Motivation des potentiellen Befragten: Auch ohne die Verwendung materieller Anreize sollte er das Gefühl erhalten, einen wertvollen Beitrag geleistet zu haben. Abhängig vom Thema kann es auch sinnvoll sein, dem Befragten Zugang zu den wichtigsten Ergebnissen der Befragung zu geben: Durch das Nennen eines Links, unter dem die Ergebnisse abrufbar sind, oder auch durch die Möglichkeit der Angabe einer Mail-Adresse, an die die Ergebnisse zugeschickt werden.

Insbesondere bei schriftlichen Fragebögen ist es notwendig, explizit auf verwendete Konventionen hinzuweisen, die nicht selbsterklärend sind. Da der Fragebogenablauf bei Online-Befragungen durch die Programmlogik gesteuert wird, dürfte dies hier allerdings selten notwendig sein.

¹⁴ Mit Hilfe von JavaScript lassen sich auch raffiniertere Verfahren zur unbemerkten Identifikation realisieren.

¹⁵ Ein Online-Shop könnte z.B. Cookies verwenden, um das Einkaufsverhalten einer Person mit ihren Antworten in einer durch den Shop durchgeführten, scheinbar „anonymen“ Umfrage in Beziehung zu setzen.

3.1.2. Verabschiedung

Die Verabschiedung sollte mindestens ein freundliches „Danke“ enthalten. Je nach Befragungssituation kann auch ein „De-Briefing“ notwendig sein, dass den Befragten über den wahren Zweck der Befragung aufklärt.

Hier können auch zusätzliche Funktionen untergebracht werden, z.B. die Möglichkeit der Angabe der Adresse oder die Ausgabe eines Codes, falls Incentives verwendet werden, bzw. ein Link auf die Befragungsergebnisse.

3.1.3. Hauptteil

Der Hauptteil des Fragebogens sollte in sinnvolle Abschnitte aufgeteilt werden. Dabei ist darauf zu achten, dass thematisch zusammengehörende Fragen zu einem Abschnitt oder einer Fragesequenz zusammengefasst werden. Bei der Abfolge der Fragesequenzen sind eine Reihe von Faktoren zu berücksichtigen:

- Zu Beginn sollten keine „kritischen“ oder komplizierten Fragen gestellt werden. Hier muss der Befragte zunächst an die Interviewsituation gewöhnt werden. Dazu benutzt man häufig "Eisbrecherfragen", die zwar einen thematischen Zusammenhang zur Befragung aufweisen, aber einen eher "lockeren" Charakter aufweisen.
- Gegen Ende der Befragung sind komplizierte Fragen und zentrale Fragestellungen zu vermeiden, besonders bei längeren Befragungen, bei denen beim Befragten bereits "Ermüdungseffekte" auftreten können.
- Es ist darauf zu achten, dass in früheren Sequenzen keine Vorinformationen für spätere stecken.

Bsp.

Die Grobstruktur eines typischen Fragebogens zur Mediennutzung könnte so aussehen:

1. Einleitung
2. Generelle Mediennutzung
3. Spezifische Fragen zur Bewertung einzelner Medienangebote
4. Soziodemographie

3.2. Mikrostruktur

3.2.1. Struktur von Fragesequenzen

Wenn in einer Fragesequenz Fragen zum gleichen Sachverhalt auftauchen, die einen unterschiedlichen Grad an Abstraktion aufweisen, muss festgelegt werden, in welcher Reihenfolge diese Fragen gestellt werden.

- Vom Abstrakten/Allgemeinen zum Speziellen/Konkreten: Der Befragte wird erst an ein Thema herangeführt, erst zum Schluss werden konkrete, spezifische Fragen gestellt. Der Vorteil dieser Herangehensweise liegt darin, dass der Befragte kognitiv auf ein Thema eingestimmt wird, bevor spezifische Fragen gestellt werden. Unvermittelte konkrete Fragen nach einem Themenwechsel können verwirrend oder disorientierend wirken. Ein potentieller Nachteil: Die allgemeinen Fragen können Voreingenommenheit produzieren.
- Vom Konkreten zum Abstrakten: Vor der Diskussion abstrakter Konstrukte werden zunächst konkrete Beispiele angeführt. Dies ist zu empfehlen, wenn abstrakte Urteile abgegeben werden sollen – die zuvor gestellten konkreten Beispiele führen zur Aktivierung kognitiver Schemata, die dann zur Beurteilung herangezogen werden. Natürlich liegt hier die Gefahr der bewussten oder unbewussten Manipulation durch die Auswahl der konkreten Beispiele nahe.
- Kombination der beiden Vorgehensweisen: Allgemeine Fragen zur Einstimmung, konkrete Fragen, um die Beschäftigung mit dem Thema anzuregen, abstrakte Urteile.

Bsp.

- 1 a) Wie wichtig ist für Sie das Informationsangebot im Fernsehen?
- 1 b) Wie häufig sehen Sie Nachrichtensendungen im Fernsehen?
- 2) Bitte beurteilen Sie die folgenden Nachrichtensendungen bezüglich ihrer Qualität, wenn Sie sie zumindest mehrmals die Woche ansehen: ...
- 3) Wie zufrieden sind sie generell mit der Berichterstattung in Fernsehnachrichten?

Durch die Fragen 1 a und b wird der Befragte an den Themenkomplex herangeführt. Ohne diese Einleitungsfragen würde ihm der Fragekomplex 2 evtl. als abrupter Themenwechsel erscheinen. Die Fragen zu den Nachrichtensendungen regen ihn an, über seine Einschätzung von Nachrichtensendungen anhand konkreter Beispiele nachzudenken und tragen so dazu bei, sein Urteil in Frage 3 zu fundieren.

3.2.2. Fehlerquellen

Die Befragung stellt für den Befragten eine zeitlich zusammenhängende Situation dar. Die sequentielle Abfolge von Fragen kann, wie in 3.2 dargestellt, dazu beitragen, der Befragung subjektiven Sinn zu verleihen, und dem Befragten helfen, fundierte Antworten zu geben. Das ständig vorhandene Beeinflussungspotential der Fragen untereinander kann aber auch negative Konsequenzen haben:

- **Ausstrahlungseffekte** – eine Frage beeinflusst eine Folgefrage:
 - durch das Schaffen einer emotionalen Haltung zum Objekt der Befragung
 - durch das Preisgeben von Vorinformationen
- **Sequenzeffekte** – die Beantwortung aufeinander folgender Fragen weist Muster auf, die durch externe Faktoren beeinflusst werden:
 - Tendenz zur Mitte
 - Systematischer Bias: Tendenz zur Auswahl der linken (erstgenannten) Alternative
 - Bei der Auswahl von Items aus einer Liste: Unterschiedliche Wahrscheinlichkeit der Auswahl je nach Platzierung („Primacy-Effect“)

Ausstrahlungseffekte können durch Sorgfalt bei der Platzierung von Fragen vermieden werden. Zur Reduzierung von Sequenzeffekten kann man unterschiedliche Techniken heranziehen:

- Tendenz zur Mitte: Vermeiden einer „neutralen“ Antwortmöglichkeit (gerade Anzahl von Skalenpunkten)
- Systematischer Bias: Wechsel der Polarisierung von Items in Listen.
- Primacy-Effekt: Permutation der Anordnung der Items in Itemlisten

3.2.3. Filter

In vielen Befragungssituationen ist es notwendig, Teile der Befragung für verschiedene Gruppen von Befragten zu variieren, da eine Reihe von Fragen nur für eine Teilgruppe sinnvoll gestellt werden kann. Fragt man z.B. nach der Nutzung und Bewertung von Fernsehprogrammen, dann können Fragen zur Bewertung natürlich nur von Personen beantwortet werden, die das entsprechende Programm auch kennen und nutzen.

Filter können in unterschiedlicher Form zum Einsatz kommen. Wir können zunächst proaktive von retroaktive Filtern unterscheiden:

Proaktive Filter: Wenn Frage a mit „x“ beantwortet wurde, springe zu Frage n , d.h. eine Abfrage erfolgt nach Frage a und führt gegebenenfalls zum Überspringen eines ganzen Blocks.

Proaktive Filter lassen sich weiterhin in Filter mit einfachen und multiplen Verzweigungen unterscheiden.

Bsp:

Filter mit einfacher Verzweigung:

Haben Sie das TV-Programm *Vox* schon einmal angesehen?

ja? \Rightarrow Fragen zur Nutzung von *Vox*

nein? \Rightarrow überspringe Fragen zur Nutzung von *Vox*

Filter mit multipler Verzweigung:

Wie häufig nutzen Sie das Internet?

täglich? \Rightarrow Fragen zur täglichen Nutzung

mehrmals pro Woche? \Rightarrow Frage zur wöchentlichen Nutzung

seltener oder nie? \Rightarrow überspringe Fragen zur Internetnutzung

Retroaktive Filter: Vor der Präsentation einer Frage wird überprüft, ob der Befragte sie sinnvoll beantworten kann. Retroaktive Filter sind im Rahmen von Fragen mit Mehrfachantworten notwendig, wenn z.B. zunächst für eine Reihe von Medien die Bekanntheit abgefragt wird, danach nur für die bekannten Me-

dien ihre Nutzung erhoben, nur für die genutzten Medien Fragen zur Bewertung gestellt werden¹⁶.

Einsatz von Filtern in der mündlichen - nicht computergestützten - Befragung

In der mündlichen Befragung steuert ein geschulter Interviewer den korrekten Einsatz von Filtern. Im Falle eines proaktiven Filters folgt der Interviewer der Filteranweisung in Abhängigkeit vom Antwortverhalten. Die Handhabung von retroaktiven Filtern ist problematischer, denn der Interviewer muss hier auf Informationen aus vorherigen Fragen zurück greifen: Wenn in einem Fragenkomplex zunächst die Bekanntheit von Medien abgefragt wird, und in einem späteren die Nutzung der bekannten Medien, darf der Interviewer nur noch auf die Medien eingehen, die der Befragte als „bekannt“ genannt hat. Daraus resultieren besondere Anforderungen und Einschränkungen für die Fragebogengestaltung – die ursprünglichen Filterfragen müssen entweder separat erfasst werden, oder für den Interviewer leicht zugänglich sein¹⁷.

Filter in der schriftlichen Befragung

In der klassischen Pencil-and-Paper-Befragung muss der Befragte selbst die Filterführung übernehmen. Daraus resultieren zwei Nachteile:

1. Vom Befragten wird eine zusätzliche Leistung verlangt, die die Beantwortung kompliziert. Komplexe Filterführung ist hier nicht möglich, da vom Befragten der Aufwand zum Lernen des Filters nicht verlangt werden kann.
2. Der Befragte sieht Teile des Fragebogens, die nicht für ihn bestimmt sind und die ihn von der eigentlichen Befragung ablenken können.

In der schriftlichen Befragung wird man daher auf Filterführung entweder ganz verzichten, oder nur sehr einfache Filter einsetzen.

¹⁶ Ein klassisches Beispiel ist der Fragekomplex zur Nutzung verschiedener Medien in der Media Analyse (MA): Für jedes Medium wird zunächst gefragt, ob der Titel bekannt ist, für die bekannten Titel, ob man eine der letzten 12 Ausgaben genutzt hat, für die genutzten Titel wird abschließend die Nutzungshäufigkeit ermittelt.

¹⁷ In den persönlich-mündlichen Befragungen der MA wird dieses Problem mit der Hilfe von Vorlagekarten gelöst, die der Befragte nach bekannt/nicht bekannt, genutzt/nicht genutzt sortiert.

Filter in der computergestützten Befragung

In jeder Form computergestützter Befragung kann die Filterführung programmiert werden – Fehler durch den Interviewer oder, bei selbstausgefüllten Fragebögen, durch den Befragten, sind so ausgeschlossen.

Im Falle von proaktiven Filtern wird die Antwort auf die Filterfrage vom Programm ausgewertet, das in Abhängigkeit von dieser Antwort die entsprechende nächste Frage aufruft. Auch retroaktive Filter in der Folge von Mehrfachantworten sind leicht zu realisieren. Die Informationen, die über die Präsentation einer Frage entscheiden, können

- in einem Cookie gespeichert werden, das vor dem Anzeigen einer Frage abgerufen wird;
- in der Datenbank gespeichert werden;
- über die einzelnen Seiten hinweg übergeben werden.

Vor dem Aufruf einer Frage wird zunächst eine Bedingung überprüft:

- Trifft die Filterfrage zu, präsentiere die entsprechende Frage;
- trifft sie nicht zu, überprüfe die nächste Bedingung.

Gerade bei komplexen Fragebögen ist die Möglichkeit der programmierten Filterführung ein entscheidender Vorteil. Darüber hinaus lassen sich hier auch Filter beliebiger Komplexität realisieren, deren Einsatz für einen Interviewer ohne Computerunterstützung nicht in Frage kommen würde, da zu viel Information in zu kurzer Zeit verarbeitet werden müsste, z.B. Filter der Art „Die folgende Frage soll nur weiblichen Befragten über 30 Jahren gestellt werden, die mehrmals in der Woche eine Nachrichtensendung eines privaten Fernsehsenders anschauen.“

3.2.4. Eingabe- und Konsistenzchecks

Neben den Fehlern, die durch Fehler bei der Fragebogenkonzeption und –konstruktion entstehen, gibt es Fehler, die aus der Interaktion während der Befragung entstehen:

- Missverstehen der Frage durch den Befragten
- Missverstehen der Antwort durch den Interviewer

- Falsches Erinnern oder Lügen
- Falscher Eintrag durch den Interviewer

In einer mündlichen Befragung können Fehler z.T. korrigiert werden – wenn der Interviewer bemerkt, dass die Antwort sich offensichtlich nicht auf die Frage bezieht, oder wenn sie einen zulässigen Rahmen verlässt (z.B. wenn das angegebene Geburtsdatum eindeutig nicht zum Alter des Befragten passt, o.ä.).

Kontrollmöglichkeiten stehen in schriftlichen Befragungen nicht zur Verfügung, die Datenqualität ist hier mit von der Sorgfalt des Befragten abhängig. Bei CASI-Verfahren lässt sich eine Vielzahl von Fehleingaben aber durch die verwendete Software abfangen (vgl. 6.2).

3.3. Fragebogenaufbau in schriftlichen Befragungen

Prinzipiell gelten für Fragebogen die gleichen Regeln, unabhängig davon, ob er in einem mündlichen Interview verwendet wird, oder schriftlich ausgefüllt. Die andere Modalität der Darstellung führt aber zu Unterschieden in der Art der Strukturierung. In einer mündlichen Befragung erscheint der Fragebogen für den Befragten als eine sequentielle Anordnung von Fragen. Strukturierungsmittel sind hier ausschließlich verbaler Natur, etwa in der Form "Wir möchten jetzt noch etwas über ihre Mediennutzung erfahren...".

In der schriftlichen Befragung wird der Interviewte mit einer räumlichen Gliederung konfrontiert. Daher muss besonderer Wert auf das Layout gelegt werden, dem die Aufgabe der Strukturierung zukommt. Verbale Überleitungen können auch hier eingesetzt werden – da der Befragte bei einem "Pencil&Paper"-Fragebogen den Überblick über den Frageverlauf hat, kommt ihnen eine weniger große Bedeutung zu.

3.3.1. Fragebogenaufteilung: CASI

Im Gegensatz zu einer „Pencil&Paper“-Befragung kennt der Online-Fragebogen keine festen Seiteneinteilungen. Eine HTML-"Seite" kann so groß sein, wie der Fragebogendesigner es wünscht (allerdings ist es nicht sinnvoll, diese "Freiheit" auszunutzen). Der Fragebogendesigner kann hier auch frei entscheiden, wieviel Einblick der Befragte in den Verlauf des Interviews bekommen soll. Denkbar sind:

1. Komplette Übersicht über den gesamten Fragebogen, die Fragen können (insoweit die Logik des Frageablaufs es gestattet) in beliebiger Reihenfolge ausgefüllt werden.
2. Alle Fragen können in einer Preview angesehen werden, das Ausfüllen erfolgt aber in einer festgelegten Reihenfolge.
3. Der Befragte erfährt am Anfang einer Fragesequenz, um welches Thema es bei den nächsten Fragen gehen wird.
4. Der Befragte weiß zu jedem Zeitpunkt, an welcher Stelle im Fragebogen er sich befindet und kann die Zeit, die noch für die Beantwortung notwendig ist, abschätzen (bei Fragebögen mit vielen Filtern ist diese Angabe natürlich nur eine grobe Schätzung.)

Wieviel Informationen der Fragebogendesigner preisgeben will, hängt primär vom Forschungsinteresse ab. Die Entscheidung hängt auch mit der Seitenverteilung zusammen, also damit, ob der Fragebogen auf einer Seite oder mehreren Seiten präsentiert wird. Im letzteren Fall sollte man zumindest für ein Minimum an Transparenz für den Befragten sorgen (Punkt 3 und 4).

- Folgende Formen der Verteilung der Seiten sind möglich:
- Der gesamte Fragebogen ist in einem Formular untergebracht.
- Jede Seite enthält mehr oder weniger lange Fragesequenzen
- Eine Frage pro Seite
- Die Verteilungsformen haben jeweils spezifische Vor- und Nachteile.

Einseitige Fragebögen

- + Einfach zu programmieren
- nur geeignet für sehr kurze Fragebögen ohne Filterführung
- unübersichtlich
- Befragter kann Fragen in beliebiger Reihenfolge auswählen, „Verstecken“ von Informationen nicht möglich.
- Befragungsabbruch führt zu Verlust *aller* Informationen.

Sequenzen auf Seiten verteilt

- + Filterführung möglich
- + Reihenfolge der Fragesequenzen ist festgelegt

- + bessere Übersichtlichkeit, da kürzere Seiten
- kein Verstecken von Informationen innerhalb von Fragesequenzen
- problematische Rückkopplung bei fehlenden Antworten
- falls Fragebogen abgebrochen wird, ist zwar ein Teil der Daten vorhanden, der genaue Abbruch lässt sich aber nicht feststellen.

Eine Frage pro Seite

- + präzise Filterführung
- + einfache Kontrollmöglichkeit, welche Fragen beantwortet werden müssen.
- + der genaue Zeitpunkt eines Abbruchs ist feststellbar
- + Die Frageabfolge kann (z.B. in Itemlisten) automatisch variiert werden.
- hoher Programmieraufwand
- Viele Nutzer-Aktionen notwendig, insbesondere bei längeren Item-Listen ermüdend.

Variation der Verteilung

In der Regel ist es sinnvoll, die Fragenanzahl zu variieren, ausgehend von einem Modell, auf dem pro Seite nur eine oder wenige Fragen vorkommen:

- Zusammengehörende Fragen werden auf einer Seite plaziert.
- Itemlisten werden auf einer Seite zusammengefasst.

4. Stichprobenziehung

4.1. Zufallsstichprobe

In jeder empirischen Untersuchung wird der Versuch unternommen, Aussagen über einen Realitätsausschnitt zu ziehen. In der Sozialwissenschaft ist es nur selten möglich, bei einer Untersuchung den gewünschten Ausschnitt als Ganzes zum Gegenstand der Untersuchung zu machen.

Mit Hilfe von Stichprobenverfahren und der schließenden Statistik ist es möglich, auf der Basis von Ausschnitten der Grundgesamtheit gefundene Zusammenhänge im Rahmen probabilistischer Aussagen zu verallgemeinern.

Ausgangspunkt für ein derartiges Verfahren ist eine Zufallsstichprobe. Bei einer Zufallsstichprobe hat jedes Element der Grundgesamtheit die gleiche Chance, in die Stichprobe aufgenommen zu werden.

4.2. Mehrstufige Auswahlverfahren

Aus forschungspraktischen Gründen ist es zuweilen notwendig, das Stichprobenverfahren in mehreren Phasen durchzuführen. Dabei sollte das Prinzip der zufälligen Auswahl aber beibehalten werden.

4.2.1. Geschichtete Stichproben

Bei geschichteten oder stratifizierten Samples wird die Stichprobe in verschiedene Segmente oder Schichten eingeteilt. Die Größe der Segmente in der Population muss dabei bekannt sein, die Verteilung im Sample der Populationsverteilung entsprechen. Innerhalb der Schichten wird eine Zufallsauswahl durchgeführt. Dieses Verfahren führt bezüglich der Schichtungskriterien zu einer besseren Repräsentativität, da die Abweichung der Stichprobenparameter vom Populationsparameter hier kontrolliert wird. Angewandt wird das Verfahren unter anderem, um sicherzustellen, dass kleine – oder schwer zu erfassende – Bevölkerungssegmente in einer Stichprobe ausreichend repräsentiert sind.

Quotenstichproben

Quotenstichproben verfolgen einen ähnlichen segmentierenden Ansatz wie geschichtete Stichproben. Auch hier wird von einer Reihe von Merkmalen ausgegangen, deren Verteilung in der Population bekannt sind. Der Unterschied findet sich in der Stichprobenziehung: Hier erhält jeder Interviewer eine Quoten-

vorgabe, die er zu erfüllen hat – ohne dass näher bestimmt wird, wie er dies tut. Dabei wird von der Überlegung ausgegangen, dass durch die Quotierung und die Heterogenität der Interviewer Verzerrungen ausgeschlossen werden können. Dieses Verfahren, das in größerem Umfang in Deutschland nur vom *Institut für Demoskopie* in Allensbach eingesetzt wird, ist in der Umsetzung billiger als Zufallsstichproben. Allerdings lassen sich auf über Quotenverfahren gewonnene Stichproben streng genommen keine Verfahren der Inferenzstatistik anwenden.

4.2.2. Clusterstichproben

Cluster- oder Klumpenstichproben werden angewendet, wenn Umfragen in ausgedehnten Gebieten durchgeführt werden sollen. Insbesondere bei persönlich-mündlicher Befragung, die die Anwesenheit des Interviewers voraussetzt, entstehen beträchtliche Kosten durch die Anreise. Um diese Kosten zu reduzieren, wird das Gebiet, in dem die Population wohnt (z.B. das Bundesland Bayern) zunächst in Teilgebiete aufgeteilt, in denen jeweils die gleiche Anzahl Personen wohnen – dazu sind z.B. Wahlkreise gut geeignet. Zunächst wird aus diesen Teilgebieten eine Stichprobe gezogen. Die Ziehung der Personenstichprobe erfolgt dann nur innerhalb der ausgewählten *Sample Points*. Dabei bleibt das Kriterium der Zufallsauswahl erhalten, die Reisekosten werden minimiert. In der Praxis erfolgt die Auswahl der *Sample Points* häufig mehrstufig.

4.2.3. Haushalts- und Personenstichproben

Die meisten Stichprobenverfahren führen zunächst zur Auswahl von Haushalten (Adressen oder Telefonnummern), nicht Personen. Für die Befragung sind aber meistens Personenstichproben relevant. Bei der Transformation der ursprünglichen Haushaltsstichprobe in eine Personenstichprobe ist folgendes zu beachten:

- Es ist nicht ausreichend, die erste kontaktierte Person im Haushalt zu befragen (außer, es handelt sich um einen Ein-Personen-Haushalt), denn dadurch können systematische Verzerrungen auftreten.
- Je größer der Haushalt, desto geringer die Chancen für das einzelne Haushaltsmitglied, in die Stichprobe – die ja zunächst eine Auswahl auf Haushaltsebene trifft – zu geraten. Dies widerspricht der Anforderung

an eine Zufallsstichprobe, nach der jedes Element die gleiche Chance zur Auswahl haben muss. Ohne Korrektur wird eine Verzerrung in die Stichprobe eingeführt. Eine Korrekturmöglichkeit ist die nachträgliche Gewichtung, die allerdings aus statistischer Perspektive problematisch ist.

4.3. „Aktives“ und „passives“ Sampling

4.3.1. Der mehrphasige Charakter einer Stichprobenziehung

Die Stichprobenziehung erfordert in der Regel ein aktives Vorgehen des Forschers: In der Umfrageforschung kann die Zufallsstichprobe die Elemente nicht direkt produzieren, sondern lediglich Platzhalter – z.B. Telefonnummern, Post- oder eMail-Adressen. Zum Zeitpunkt der Stichprobenziehung ist der Befragte nicht in den Prozess der Auswahl involviert.

Gegenüber dem idealen Urnenmodell ist das Sampling in der Umfrageforschung in (mindestens) zwei Schritte aufgeteilt: Die Stichprobenziehung der Platzhalter und die Kontaktaufnahme mit den „Elementen“ der Stichprobe. Auch mit der Kontaktaufnahme ist das Sampling noch nicht abgeschlossen, denn der potentielle Teilnehmer muss mit mehr oder weniger Aufwand überredet werden, an der Befragung teilzunehmen.

Ein aktives Sampling ist auch im Rahmen von Online-Befragungen möglich. Z.B. kann die Kontaktaufnahme über ein telefonisches Screening auf der Basis einer Zufallsstichprobe von Telefonnummern erfolgen, um die Zugehörigkeit der Kontaktpersonen zur „Online-Population“ zu etablieren.

4.3.2. E-Mail-Stichproben

Da es keine Verzeichnisse von Mail-Adressen gibt, die in der Qualität mit Telefonbüchern vergleichbar wäre, ist ein Versuch, mit Hilfe von Mail-Kontaktierung eine „repräsentative“ Stichprobe der „Online-Population“ zu ziehen, zum Scheitern verurteilt. Das Problem ist weniger gravierend, wenn die Population eingegrenzt ist:

- wenn z.B. eine Online-Community eine Befragung unter ihren Mitgliedern durchführen will, deren Mail-Adressen bei der Registrierung angegeben werden muss.

- wenn eine Firma oder eine Organisation, deren Mitglieder aktive Mail-Nutzer sind, eine Befragung durchführen möchte.

Auch in diesem Fall kann die Qualität der Stichprobe angezweifelt werden, denn die Nutzung der Mail-Adressen kann in der ausgewählten Personen-Gruppe sehr unterschiedlich verteilt sein. Manche Personen nutzen ihre Accounts nur selten, zum Teil geben Nutzer auch absichtlich selten genutzte e-Mail-Adressen an, um nicht mit Werbemails (oder Umfragen) belästigt zu werden.

Die Spam-Problematik

Die Abgrenzung zu Werbemails („Spam“) stellt ein zentrales Problem bei der Kontaktierung per E-Mail dar. Nicht zuletzt aufgrund der Versuche von Massen-Mailern, Werbung hinter vorgeblichen Umfragen zu verstecken besteht ein gewisser Vorbehalt, auf Umfragen zu antworten. Hier ist es wichtig, die Legitimität der Umfrage deutlich zu machen, einerseits durch einen Hinweis in der Betreffzeile, der Nennung eines Ansprechpartners, der möglichst auch telefonisch zu erreichen ist und der Verwendung von „legitimen“ Mailadressen und Servernamen. Eine Rekrutierungsmail für eine Online-Umfrage der Uni München, die von einer Adresse „honey69@hotmail.com“ abgesendet wird, und deren URL „www.p0rnstar.com/suckers/survey.php“ lautet, wird Misstrauen hervorrufen.

Eine neue Problematik bei der Stichprobenziehung per Mail ist die technische Differenzierung der Rekrutierungsmail von „Spam“. Prinzipiell empfiehlt es sich nicht, eine Mail an eine längere Liste mit Adressaten zu senden – zum einen empfinden die Empfänger diese Form der Ansprache als unpersönlich, zum anderen sortieren viele Spam-Filter Mails mit umfangreichen Empfängerlisten schnell aus. Leider haben größere Anbieter von Mail-Diensten ihre Spam-Filter dahingehend erweitert, dass Mails mit ähnlichem oder gleichem Inhalt, die zeitnah an viele Nutzer des Mail-Dienstes verschickt werden, ebenfalls in die Kategorie Spam eingeordnet werden. Leider ist das technische Profil des massenhaften Versandes von Rekrutierungsmails für Umfragen mit dem von Spam weitgehend identisch, die Online-Forschung gerät hier ungewollt zwischen die Fronten im Konflikt zwischen Spammern und Spam-Blockern. Umso

wichtiger ist es, typische Kennzeichen von Spam-Mail im Anschreiben zu vermeiden. Dazu gehören:

- Verwendung von HTML oder Rich Text zur Formatierung von Mails
- Wörter oder Textabschnitte in Versalien in der Subjectzeile oder im Text selbst.

Vor dem eigentlichen Versand ist ein Test zu empfehlen, bei dem die Mail gängige Spamfilter durchläuft.

4.3.3. Alternativen der aktiven Stichprobenziehung und Kontaktaufnahme

Aktive Verfahren: Zufällige Auswahl und gezielte Kontaktaufnahme mit der Person, die befragt werden soll.

- Haushaltsstichprobe auf der Basis von Random-Route-Verfahren mit anschließender Zufallsauswahl der zu befragenden Person.
 - Vorteile: Gute Stichprobenqualität, erprobtes Verfahren
 - Nachteile: Hoher Streuverlust, sehr hohe Kosten. Wenn bereits ein persönlicher Kontakt mit hohem Aufwand erfolgt ist, ist es in der Regel kaum sinnvoll, eine Online-Befragung anzuschließen. Allerdings kann es durchaus sinnvoll sein, auch hier computergestützte Befragungstechniken (CAPI) einzusetzen.
- Telefonstichprobe: Hier wird inzwischen meist „Random Last Digit Dialing“ eingesetzt, das auf der zufälligen Veränderung der letzten Ziffern von aus dem Telefonbuch gewonnenen Nummern basiert¹⁸.
 - Vorteile: Gute Stichprobenqualität, erprobtes Verfahren, zum großen Teil automatisierbar, geringerer Zeitaufwand und Kosten als persönlicher Besuch der Haushalte.
 - Nachteile: Hoher Streuverlust im Bezug auf die angestrebte Population. Für eine One-Shot-Befragung dürfte der Aufwand zu hoch sein. Allerdings bietet die Telefonstichprobe eine noch kostengünstige Möglichkeit, Teilnehmer für ein Online-Panel anzuwerben.

¹⁸ Eine Generierung von kompletten Telefonnummern auf Zufallsbasis würde zu sehr vielen nicht existenten Nummern führen, da der Aufbau der Nummern bestimmten Regeln folgt. Eine Stichprobe aus Telefonbüchern kann die zunehmende Zahl der nicht gelisteten Nummern nicht erfassen.

- E-Mail-Stichprobe: Man kann davon ausgehen, dass fast jedes Mitglied der „Online-Population“ eine Mail-Adresse besitzt. Theoretisch wäre die Ziehung einer Stichprobe aus dieser Grundgesamtheit und die Kontaktaufnahme per Mail das idealtypische Stichprobenverfahren.
 - Vorteile: Geringe Streuverluste, kostengünstig, schnell.
 - Nachteile: Aus mehreren Gründen ist eine Mail-Stichprobe meist von geringer Qualität.
 - i. Es existiert keine Aufstellung von Mail-Adressen, mit deren Hilfe die „Online-Population“ als Ganzes beschrieben werden kann und die zum Zweck der Kontaktaufnahme genutzt werden könnte. Derartige Verzeichnisse existieren nur für wenige Teil-Populationen (Firmen, Organisationen, Vereine).
 - ii. Es existiert keine Eins-zu-eins-Zuordnung zwischen Mail-Adressen und Personen. Manche Familien teilen sich einen eMail-Account, andererseits haben viele Personen mehrere Mail-Adressen (und darum bei einer Zufallsauswahl von Adressen größere Auswahlchancen).
 - iii. Die Nutzungsmuster von eMail sind sehr heterogen – während man die meisten Leute innerhalb von einer Woche persönlich oder per Telefon zu Hause erreichen kann, werden manche eMail-Accounts nur selten abgefragt. Personen, die häufiger abfragen, haben eine größere Auswahlwahrscheinlichkeit.
 - iv. Die Kontaktqualität ist nicht hoch einzuschätzen – Anfragen zur Teilnahme an einer Umfrage können leicht als „Spam“ oder als illegitime Marketing- oder Werbemaßnahme eingeschätzt werden¹⁹. Es besteht kaum eine Möglichkeit, die angesprochene Person zu überzeugen, warum eine Teilnahme sinnvoll sein kann²⁰.

Bei den aktiven Verfahren hat man also die Wahl zwischen aufwendigen und teuren Vorgehensweisen, oder einem preisgünstigen Verfahren mit einer Reihe

¹⁹ Leider besteht diese Einschätzung häufig zu Recht, nicht zuletzt dank einer völlig unzureichenden, verbraucherfeindlichen rechtliche Regelung von Werbemails.

²⁰ Die einfache und kostengünstige Umsetzung einer Mail-Stichprobe erweist sich hier als Bumerang: Denn der Befragte weiß natürlich, dass im Gegensatz zur persönlichen oder telefonischen Kontaktaufnahme beim Versenden einer eMail kaum Aufwand notwendig ist – wenn man sich so wenig um ihn bemüht, warum sollte er dann Anstrengungen unternehmen, um den Fragebogen auszufüllen?

gravierender Nachteile. Man wird das Verfahren wählen, das eine akzeptable Relation zwischen notwendiger Qualität und Aufwand bietet, abhängig von der Fragestellung. Auch eine E-Mail-Stichprobe ist einem passiven Stichprobenverfahren (vgl. 4.3.5) überlegen und kann z.B. im Rahmen einer Mitarbeiterbefragung gute Ergebnisse liefern.

4.3.4. Anwendbarkeit mehrstufiger Verfahren für die Online-Befragung

Insofern man sich im Rahmen von Online-Befragungen traditioneller Stichprobenverfahren bedient, sind die bisherigen Überlegungen zu den Vorteilen mehrstufiger Verfahren auch hier relevant. Anders sieht es aus, wenn man sich einer E-Mail-Stichprobenziehung bedient:

- Wenn man nicht davon ausgeht, dass die Mitglieder größerer Haushalte schlechter über Mail zu erreichen sind, stellt sich das Problem der Transformation einer Haushalts- in eine Personenstichprobe geht. Da man in jedem Fall von einer Online-Population (vgl. 4.3.8) ausgehen muss, lässt sich das Konzept einer Haushaltsstichprobe kaum auf E-Mail-Sampling übertragen.
- Da weder bei der Kontaktaufnahme noch der Befragung Kosten für Anreise oder Ferngespräche entstehen, ist die Notwendigkeit einer Clusterstichprobe nicht gegeben.
- Eine geschichtete Stichprobe lässt sich theoretisch auch bei einer E-Mail-Stichprobe einsetzen und könnte dazu beitragen, ihre Qualität zu verbessern. Dabei müsste die Kontaktierung mehrstufig erfolgen, wobei ein ständiger Abgleich von Screening-Fragen zur Etablierung der Zugehörigkeit zu dem gewünschten Segment mit den gespeicherten Daten erfolgen müsste.

4.3.5. Passive Verfahren

Neben der klassischen aktiven Stichprobenziehung hat sich im Web auch eine Form der passiven Stichprobenziehung verbreitet. Der Vorläufer dieser Variante des Sampling sind Umfragen in Beilagen zu Printprodukten. Hier erfolgt ein zufälliger Kontakt mit einem Hinweis auf eine Befragung (oder auch den Fragebogen selbst), die Person entscheidet selbst, ob sie sich angesprochen fühlt

und an der Befragung teilnimmt. Die Problematik dieser Stichprobenziehung liegt darin, dass die Rahmenbedingungen, die letztlich zu der Teilnahme führen, unkontrollierbar sind.

Häufig verwendete passive Verfahren sind:

- Platzierung auf von der Population hoch frequentierten Websites, z.B. Portalen
 - Vorteile: Relativ geringe Kosten
 - Nachteile: Noch stärker als bei der Mail-Stichprobe steuert das Verhalten der potentiellen Befragten die Stichprobe. Selbstselektion ist nicht ein Problem einer derartigen Stichprobe, sondern ihr Prinzip.
 - i. Die „Ausschöpfung“²¹ ist in der Regel auch dann äußerst gering, wenn die Zahl der Befragten hoch ist. Die Click-Through-Raten vom Banner, das den Hinweis zur Befragung enthält, bewegen sich im Promillebereich
 - ii. Wer wann und wie häufig eine Website auswählt, ist nicht allein vom Zufall gesteuert. Die Chance, Vielnutzer zu erreichen, ist höher als bei Wenignutzern. Auch das Streuen von Werbebannern über verschiedene Websites mildert das Problem nicht.
 - iii. Jeder, der den Hinweis auf die Befragung sieht, kann selbst entscheiden, ob er an der Befragung teilnehmen wird. Diese Entscheidung kann von vielen Aspekten abhängen, unter anderem vom Thema der Befragung. Damit sind systematische Verzerrungen bereit in das Auswahlverfahren eingebaut.
- Hinweise auf die Befragung in Foren, Newsgroups, Mailinglisten.
 - Vorteile: In der Regel geringe Streuverluste, wenn das Thema der Befragung mit dem Thema des Forums korrespondiert. Überdurchschnittlich hohe Motivation der Befragten.
 - Nachteile: In stark frequentierten Foren wird der Hinweis auf die Umfrage nur geringe Aufmerksamkeit erfahren. Die Einschränkung auf ein oder weniger Foren bringt eine Verzerrung der Stichprobe mit sich, deren Parameter nicht einzuschätzen sind.

²¹ Man kann hier streng genommen nicht von „Ausschöpfung“ reden, da es keinen Bruttoansatz für die Stichprobe gibt.

Eine Befragung, die aufgrund einer passiven Stichprobe getroffen wird, ist nicht repräsentativ²². Es gibt kein Verfahren, die gewonnenen Daten so zu transformieren, dass Schlussfolgerungen auf eine wie immer geartete Population möglich sind. Jeder Autor, jedes Marktforschungsunternehmen, das derartige Stichproben benutzt um Schlüsse auf das Verhalten „der Internetnutzer“ zu ziehen, hat Probleme mit der Kompetenz oder der Aufrichtigkeit.

Sind Befragungen aufgrund derartiger passiver Samples damit sinnlos? Nicht unbedingt, denn es gibt eine Vielzahl von Fragestellungen, für die ein repräsentatives Sample gar nicht notwendig ist. Beispiele dafür sind explorative Studien, die z.B. dazu dienen, mögliche Fragestellungen zu testen, oder experimentelle Ansätze, bei denen mit randomisierten Gruppen gearbeitet wird und bei denen Interaktionen der experimentellen Faktoren mit Randbedingungen ausgeschlossen oder kontrolliert werden können. Auch bei Evaluationen von Websites durch deren Besucher wird man die Tatsache, dass die Befragung kein repräsentatives Bild der Nutzer abbildet, in der Regel hinnehmen können.

4.3.6. Vollerhebungen

Stichprobenziehung in der Umfrageforschung ist meist ein notwendiges Übel, da eine Vollerhebung zu teuer ist. Online-Umfragen sind dagegen preisgünstig zu realisieren, warum also eine Stichprobe ziehen?

Ein grundsätzliches Problem bei einer Vollerhebung besteht darin, dass man die die Population nicht nur definieren muss, sondern jedes zugehörige Element definieren. Bei einer Stichprobe ist es lediglich notwendig, bei jedem Element eine Entscheidung über die Zugehörigkeit zur Population zu treffen – die Vorgehensweise entspricht einer falsifizierenden Hypothesenüberprüfung, bei einer Vollerhebung ist es darüber hinaus notwendig, die Vollständigkeit der Erhebung zu verifizieren – bei größeren Populationen ist dies sehr aufwendig.

Ein weiteres Argument gegen die Verwendung von Vollerhebung ist die Gefahr der Überforschung. Würden Kommunikationsforscher über ein Verzeichnis aller Journalisten verfügen und bei jeder Befragung ihre Population anschreiben, dürfte die Motivation zur Teilnahme an der Befragung schnell abnehmen.

²² D.h., sie könnte rein theoretisch repräsentativ sein, aber es gibt keine Möglichkeit, dies zu belegen. Ein Vergleich weniger soziodemographischer Kennwerte, der häufig als Nachweis von Repräsentativität herangezogen wird, ist tatsächlich nicht ausreichend.

Schließlich besteht bei der Verwendung von Vollerhebungen die Versuchung, den Rücklauf der Erhebung wie eine Stichprobe zu behandeln. Tatsächlich liegt hier aber das gleiche Verhältnis vor wie zwischen Netto- und Bruttoansatz einer Stichprobe. Versuche, mit Hilfe inferenzstatistischer Verfahren vom Rücklauf einer Vollerhebung auf die Population zu schließen, sind nicht legitim.

Praktikabel ist eine Vollerhebung daher auch hier nur bei einer kleinen Population, die vollständig online erreichbar ist, z.B. bei einer Mitarbeiterbefragung, einer Untersuchung eines Diskussionsforums oder einer Usenet-Newsgroup oder bei einer Studie der Akzeptanz und Nutzung eines Web-Auftritts (die Population definiert sich hier über die Nutzung der Site).

4.3.7. Online-Panels

Ein Online-Panel ist eine Gruppe von Personen, die wiederholt per Online-Umfrage befragt wird. Die Rekrutierung eines Online-Panels ist aufwendig – das Panel soll für die entsprechende Population repräsentativ sein, und gleichzeitig technisch über das Internet erreichbar sein. Dies ist natürlich eher für Populationen realisierbar, die überdurchschnittlich häufig zu den Internetnutzern gehören. Da die Panelmitglieder wiederholt befragt werden, muss man davon ausgehen, dass sie mit der Hilfe von Incentives zur permanenten Teilnahme motiviert werden. Der wiederholte Einsatz der gleichen Stichprobe birgt natürlich die Gefahr systematischer Verzerrungen. Ein einmal eingerichtetes Online-Panel hat den Vorteil, schnell und kostengünstig Umfrageergebnisse produzieren zu können.

4.3.8. Die Online-Population

Jede Population, die im Rahmen sozialwissenschaftlicher Forschung untersucht wird, kann durch eine Erhebung nicht vollständig erschlossen werden. Baut die Studie auf Haushaltsbefragungen auf, werden Insassen von Gefängnissen, Nervenheilanstalten, Krankenhäusern oder Altersheimen ebenso ausgeschlossen wie Obdachlose. Eine Telefonumfrage grenzt alle aus, die über keinen Telefonanschluss verfügen. Die Problematik der „Online-Population“ ist eine quantitative, keine qualitative – im Gegensatz zum Telefon verfügt nur ein vergleichsweise geringer Teil der Bevölkerung über einen Zugang zum Internet, und die entsprechende Gruppe entscheidet sich in mehreren Merkmalen vom Rest der

Bevölkerung²³. Darüber hinaus ist die Frage der Zugehörigkeit zur „Online-Population“ durchaus umstritten – denn „online“ zu sein – und für Umfragen erreichbar, ist eine Frage des Verhaltens, und dies variiert bei den Internetnutzern erheblich.

4.4. Problemfelder bei der Stichprobenziehung in der Online-Forschung

4.4.1. Das Stichproben-Dilemma der Online-Forschung

Die wichtigsten Einflussfaktoren, die zu Stichprobenverzerrungen im Kontext von Online-Befragungen führen, sind:

- Unterschiedliche Nutzung: Häufige Nutzer haben eine größere Chance, mit dem Fragebogen in Kontakt zu kommen.
- Unterschiedliches Interesse: Personen mit einem größeren Interesse an der Befragung werden mit einer höheren Wahrscheinlichkeit teilnehmen.
- Zeitbudget: Eine Person mit viel Zeit wird eher bereit sein, den Fragebogen auszufüllen.
- Kostengründe: Online-Befragungen sind die einzigen, die dem Befragten Kosten verursachen können. Nutzer, die über eine Flatrate verfügen, oder an ihrem Arbeits- oder Studienplatz das Web nutzen, bei denen also keine Kosten anfallen, werden sich eher Zeit nehmen.
- Willkürliche Nutzung: Ein Nutzer kann, ohne soziale Kontrolle oder Sanktionen fürchten zu müssen, einen Fragebogen inkonsistent, falsch, oder mehrfach ausfüllen.

Lösungsansätze

Bei einer passiven Stichprobe kann eine breite Streuung von Hinweisen auf den Fragebogen und eine längerfristige Plazierung kann die Unterschiede in der Kontaktwahrscheinlichkeit nivellieren. Allerdings verursacht diese Vorgehensweise hohe Kosten. Bei einer Befragung, die sich auf die Nutzer einer einzelnen Website bezieht, ist nur die Vorgehensweise der längerfristigen Plazie-

²³ Die Situation ähnelt derjenigen in den neuen Bundesländern zu Beginn der 90er Jahre in Bezug auf Telefonumfragen. Aufgrund der niedrigen Abdeckung mit Telefonanschlüssen war eine repräsentative Befragung per Telefon hier nicht zu realisieren.

rung möglich. Bezieht sich die Befragung auf eine Personengruppe, die einen Server im Schnitt sehr häufig nutzt, sinkt das Ausmaß des Problems.

Zum Ausgleich auftretender Kosten liegt das Anbieten von Incentives nahe – z.B. in Form der Teilnahme an einem Gewinnspiel. Allerdings muss der Befragte in diesem Fall seine Anonymität aufgeben – hier führt man also eine neue Quelle für Verzerrungen ein.

Das absichtlich falsche oder inkonsistente Ausfüllen von Fragebögen lässt sich nur durch Konsistenzchecks während der Datenanalyse entdecken. Mehrfaches Ausfüllen lässt sich durch die Abspeicherung und Überprüfung der IP-Adresse des Clients abfangen – allerdings gibt es natürlich keine eindeutige Zuordnung zwischen der Adresse und dem Befragten – mehrere Personen könnten über den gleichen Client den Fragebogen ausfüllen, durch Verwendung von Proxies oder NAT (Network Address Translation) können verschiedene Clients die gleiche IP erhalten.

4.4.2. Selbstselektion bei der Stichprobenziehung

Im Gegensatz zur Stichprobenziehung bei „unbelebten“ Gegenständen – Kugeln aus Urnen, Zeitungsartikeln aus dem Archiv – stellt sich bei einer Stichprobenziehung aus einer menschlichen Population ein Rekrutierungsproblem. Im Prozess der Stichprobenziehung wird jedes Element so behandelt, als ob die Zufallsauswahl allein über die Zugehörigkeit zur Stichprobe entscheiden würde. Im Kontext einer Befragung spielt aber die Rekrutierungsphase die entscheidende Rolle über die tatsächliche Auswahl. Und in dieser Phase ist der Forscher nicht auf die Qualität seines Auswahlmodells angewiesen, sondern auf die Kooperation der potentiellen Befragten.

Die Rahmenbedingung, die für den Forscher (aus Erwägungen der Stichprobenqualität) ideal wäre, ist diejenige des Teilnahmewangs (wie im Falle staatlich angeordneter Volkszählungen). Statt Zwangsmaßnahmen steht dem Forscher nur die Möglichkeit offen, die per Stichprobe ausgewählten Personen zur Teilnahme zu überzeugen.

Die Hindernisse, die bei der Überzeugung zu überwinden sind, sind individuell verschieden. Die Bedeutung des persönlichen Kontakts bei der Rekrutierung

liegt darin, diese Unterschiede zu nivellieren. In dieser Hinsicht schneidet die Rekrutierung über computervermittelte Kommunikation schlecht ab.

- Bei der aktiven Ansprache per E-Mail kann im Gegensatz zum persönlichen oder telefonischen Kontakt nicht auf Rückfragen eingegangen werden, der Versuch einer Überzeugung oder Überredeung im Prozess der Interaktion ist nicht möglich.
- Bei passiven Verfahren – Posting in einer Newsgroup, Banner auf einer Website – ist die Entscheidung, wer befragt werden soll, dem Forscher völlig aus der Hand genommen.
- Die Faktoren, die die Selbstselektion der Befragten steuern, sind vielschichtig – auf jeden Fall unterliegen sie einer Systematik. Allgemein dürften folgende Faktoren ausschlaggebend sein:
 - Interesse am Thema der Befragung
 - Zur Verfügung stehende Zeit
 - im Falle von persönlichen Befragungen: Sympathie für den Interviewer
 - Für Online-Befragungen kommen folgende Faktoren hinzu:
 - Kosten des Online-Zugangs
 - Adäquate technische Ausstattung

4.4.3. Incentives als Mittel zur Reduktion von Selbstselektion

Die Verwendung von Incentives ist in der klassischen wie der Online-Befragung umstritten. Je mehr Aufwand eine Befragung für den Befragten bedeutet, desto eher wird man Anreize verwenden, um ihn zur Kooperation zu überreden. Als die Funkanalyse Bayern noch in Form einer Tagebuchstudie durchgeführt wurde, bei der die Teilnehmer jeden Tag detaillierte Aufzeichnungen über ihr Mediennutzungsverhalten machen mussten, erhielten sie als "Dankeschön" ein Lotterielos. In dieser Größenordnung sind Incentives unproblematisch. Ob die gängige Praxis, bei Arztbefragungen im Auftrag von Pharmaunternehmen sehr kostspielige Anreize einzusetzen, dürfte eher zweifelhaft sein. Je höher die "Entlohnung", desto größer wird die Gefahr, dass die Teilnahme nur um der materiellen Anreize willen erfolgte. Im Rahmen eines knap-

pen Forschungsbudgets sind derartige Ausgaben häufig nur schwer zu rechtfertigen.

Im Kontext von Online-Befragungen spricht ein Argument für die Verwendung von Incentives: Den Befragten entstehen zum Teil Kosten durch die Teilnahme – eine Verzerrung der Stichprobe durch die Über-Repräsentation von Personen mit kostenlosem Internetzugang (oder mit Flatrates) könnte durch materielle Anreize ausgeglichen werden.

Doch die Vergabe materieller Anreize bedingt, dass der Befragte seine Anonymität preisgibt – und sich nicht ganz sicher sein kann, ob die Angabe seiner Adresse nicht doch zusammen mit seinen Befragungsdaten gespeichert wird. Aus dieser Perspektive ist es wichtig, nach Möglichkeiten zu suchen, eine anonyme Form der Belohnung zu verwenden²⁴.

²⁴ Hier würde sich zum Beispiel die Möglichkeit anbieten, kostenpflichtige Angebote im Internet zu nutzen – z.B. der Download von Musiktiteln, ohne seine Anonymität aufgeben zu müssen.

5. Zustellung und Präsentation

5.1. Zustellung des Fragebogens

In den meisten Befragungsformen ist die Zustellung des Fragebogens ein bewusster Akt des Forschers oder seiner Mitarbeiter – in der Regel wird der Fragebogen persönlich überbracht oder mit der Post verschickt. Eine Ausnahme bilden „Supplement-Fragebögen“ in Zeitschriften, und Online-Fragebögen im Web, insofern die Stichprobenziehung passiv erfolgt. In diesem Fall ist der Fragebogen ein Anhängsel eines separat genutzten Mediums, die Entscheidung, ihm Aufmerksamkeit zu schenken stellt eine bewusste Handlung des Befragten dar.

Eine Zustellung, die nicht durch den Interviewer erfolgt, ist auf technische Hilfsmittel angewiesen, etwa die Post, oder im Falle der Online-Befragung, die Präsentation im Web. Bei einer Online-Umfrage sind daher technische Aspekte zu beachten: Zunächst muss dafür Sorge getragen werden, dass der verwendete Link korrekt wiedergegeben ist. Falls der Link nicht im Web oder einer E-Mail angeführt ist, sondern in gedruckter oder gesprochener Form übermittelt wird, ist es wichtig, eine kurze, prägnante, eindeutige URL zu verwenden: Bei der Eingabe von „www.ifkw.lmu.de/fragebogen“ sind Übermittlungs- und Eingabefehler weniger wahrscheinlich als bei einer Adresse wie „server.ifkw.uni-muenchen.de/projekte/umfragen/36876/main_1.php“.

Nach dem Aufrufen des Fragebogens muss eine dauerhafte Übertragung gewährleistet sein – eine längere Unterbrechung der Verbindung kann dazu führen, dass das Ausfüllen des Fragebogens nicht mehr fortgeführt werden kann. Darüber hinaus sind unterschiedliche Übertragungsraten zu berücksichtigen – ein Großteil der Internet-Nutzer ist nach wie vor über ein relativ langsames Telefonmodem angeschlossen. Die Präsentation von Multimedia-Material über derartig langsame Verbindungen dauert lange und dürfte zu systematischen Abbrüchen führen.

5.2. Präsentationsformen

Die Präsentation eines Fragebogens kann auditiv oder visuell erfolgen (oder in einer Kombination) – beide Formen haben Vor- und Nachteile. In den traditio-

nellen Befragungsformen ist die auditive Präsentation in der Regel *interaktiv*, die visuelle Präsentation *nicht-interaktiv*.

Interaktiv/auditiv (durch Interviewer):

- Vorteile:
 - erfordert wenig Anstrengung vom Befragten;
 - persönliche Ansprache;
 - lineare Präsentation, Kontrolle über die Informationen durch Interviewer.
- Nachteile
 - Verständnisprobleme.
 - Interviewer muss präsent sein (Kosten, Gefahr der Beeinflussung).

Nicht-interaktiv/visuell (ohne Interviewer):

- Vorteile:
 - kein Interviewer notwendig;
 - Befragter hat mehr Möglichkeiten, Verständnisprobleme durch Nachlesen zu beseitigen.
- Nachteile:
 - keine Kontrolle des Befragungsverlaufs;
 - keine Möglichkeit der persönlichen Interaktion.

Auch bei einer visuellen Präsentation kann ein Interviewer präsent sein – was einige Nachteile ausgleicht.

In CASI-Befragungen lassen sich die Defizite der nicht-interaktiven Präsentation bis zu einem gewissen Grad dadurch kompensieren, dass man Interaktionsmöglichkeiten in die Benutzerschnittstelle integriert – z.B. Nachfrage und Kontrollmöglichkeiten. Diese Kompensation betrifft aber nur die kognitiven Aspekte der Interaktion – Missverständnisse und Versäumnisse können so ausgeräumt werden, aber nicht die motivierenden sozialen Aspekte.

Trotz der durch die Befragungssoftware hergestellten Interaktionsmöglichkeiten ist der Befragte in der Online-Befragung weitgehend auf sich selbst gestellt. Im Gegensatz zur Paper&Pencil-Variante muss der Befragte hier außerdem über zumindest rudimentäre Computerkenntnisse verfügen. Bei der Program-

mierung des Fragebogens muss darauf geachtet werden, dass die Präsentation möglichst reibungslos erfolgt. Auftretende Probleme könnten Befragte mit geringen Computerkenntnissen verwirren und möglicherweise zum Abbruch bewegen. Derartige Stolperfallen sind meist technischer Natur:

- Beim Einsatz von client-seitigen Scriptsprachen – z.B. JavaScript – muss darauf geachtet werden, dass das Verhalten des Fragebogens auch dann berechenbar ist, wenn der Browser die Scriptsprache nicht unterstützt. Ein Verzicht auf die Scriptsprache darf auch keine Abweichungen in der Präsentation zur Folge haben – ansonsten gäbe es eine systematische Verzerrung, deren Ursprung in den Daten nicht nachvollziehbar ist. Der völlige Verzicht auf eine Scriptsprache ist die beste Lösung.
- Der Einsatz von Plugins ist noch problematischer. Diese werden in der Regel zur Präsentation von Multimedia-Inhalten genutzt. Fehlt das Plugin, wird der Browser dies dem Nutzer in der Regel mitteilen und zum Download auffordern. Die Nutzerreaktion in diesem Moment ist nicht abzusehen.

Das Ausführen von Scripts und der Download von Plugins ist ein sicherheitsrelevanter Akt. Viele Firmen sperren den Plugin-Download per Firewall, sicherheitsbewusste Nutzer deaktivieren Javascript in ihrem Browser. Die dadurch möglicherweise auftretenden Verzerrungen sind keineswegs trivial.

Ein Vorteil bei der Präsentation eines Online-Fragebogens liegt in der Kontrolle des Frageverlaufs durch den Forscher – diese Eigenschaft erwirbt er durch seine Zugehörigkeit zu den computergestützten Befragungsformen. Im Gegensatz zum Paper&Pencil-Fragebogen hat hier der Befragte nicht die Möglichkeit, den Fragebogen zunächst durchzublättern und Informationen einzusehen, die ihn möglicherweise beeinflussen könnten (außer, die Programmierung sieht dies explizit vor). Auch das Rückblättern und Verändern bereits eingegebener Informationen lässt sich verhindern.

5.3. Situativer Charakter der Online-Befragung

Zum Zeitpunkt der Präsentation des Fragebogens – der Durchführung der Befragung – wird für den Befragten eine neue Situation erzeugt, in der seine Handlungen in einer spezifischen Art und Weise einschränkt. Bei der Befra-

gung in der Wohnung oder dem Telefoninterview befindet sich der Befragte in einer vertrauten Umgebung in einer weitgehend vertrauten Kommunikationssituation – die durch den gesteuerten Ablauf der Befragung allerdings durch neue Interaktionsregeln modifiziert wird. Bei einer schriftlich-postalischen Befragung ist die Situation, in der der Fragebogen präsentiert wird, nicht mehr durch den Interviewer kontrollierbar – der Fragebogen kann in aller Ruhe zu Hause ausgefüllt werden, in der U-Bahn oder in der Arbeit. In noch höherem Maße gilt dies für die Online-Befragung – denn Online-Nutzung findet häufig außerhalb des häuslichen Bereichs statt – in der Ausbildungsstätte, am Arbeitsplatz, im Internet-Cafe oder unterwegs. Mit der Zunahme mobiler Internetnutzung über WLAN-Hotspots und UMTS steigt die Wahrscheinlichkeit, dass Online-Fragebögen an „ungewöhnlichen“ Orten ausgefüllt werden. Die Charakteristik des Fragebogen-Ausfüllens – mehr Aktivität als z.B. beim Zeitungslesen, aber weniger kognitiver Anspruch als Arbeit am Computer – macht ihn als Beschäftigung für Wartezeiten attraktiv.

Für den Befragten ergibt sich aus der spezifischen Zustellung und Präsentation eine ausgesprochen unverbindliche Kommunikationssituation. Damit er die naheliegende Möglichkeit des Abbruchs nicht leichtfertig nutzt, sollte folgendes beachtet werden.

- Verbindlichkeit schaffen
- kurze Fragen formulieren
- keine Monotonie aufkommen lassen

Da keine Gesprächssituation vorliegt, ist es nicht notwendig, Floskeln zu verwenden, um die für den Befragten den Eindruck eines normales Gesprächs zu erwecken. Überleitungen und ausschweifende Formulierungen erhöhen dagegen die Gefahr, dass Befragte mit knappem Zeitbudget die Befragung abbrechen.

6. Erfassen der Antworten

6.1. Datenspeicherung

Die Daten werden – nach der Fehlerprüfung – auf dem Server gespeichert. Bei Ein-Seiten-Fragebögen kann dies in einer normalen Textdatei geschehen. Bei mehrseitigen Fragebogen ist die Verwendung einer Datenbank angebracht.

Es ist sinnvoll, bei der Speicherung außer den Antworten eine Reihe zusätzlicher Informationen mit abzuspeichern.

- Zeitpunkt der Abspeicherung. Minimal sollte hier der Zeitpunkt erfasst werden, an dem die Befragung begonnen wurde, um später einen Überblick über einen zeitlichen Verlauf der Responses zu bekommen. Zeitangaben bieten darüber hinaus weitere Möglichkeiten:
 - Die zusätzliche Zeiterfassung bei der Beantwortung der letzten Frage gibt Auskunft über die Gesamtdauer der Befragung.
 - Eine separate Zeiterfassung bei jedem Speichervorgang ermöglicht eine Diagnose des Befragungsverlaufs: Wurde zur Beantwortung bestimmter Fragen sehr lange gebraucht? Unterscheiden sich Gruppen bei der Beantwortungsdauer? Führen unterschiedliche Frageformulierungen oder Präsentationstechniken zu Zeitunterschieden? Diese Diagnosemöglichkeiten helfen insbesondere bei der Diagnose von Abbruchmustern. Außerdem lässt sich überprüfen, ob Teile des Fragebogens so schnell ausgefüllt wurden, dass nicht von einer ernsthaften Beantwortung ausgegangen werden kann.
- IP-Adresse des Web-Clients. Wird ein Fragebogen innerhalb kurzer Zeit mehrfach von der gleichen IP-Adresse aus ausgefüllt, liegt der Verdacht nahe, dass diese Fragebögen nicht ernsthaft beantwortet wurden. Dieser Verdacht sollte allerdings noch durch eine Analyse der Antwortmuster bestätigt werden. Im Regelfall wird man hier nicht von einer Beeinträchtigung der Anonymität des Befragten ausgehen müssen – wenn allerdings eine Mitarbeiterbefragung vorgenommen wird, bei denen die Mitarbeiter von Rechnern mit festen IP-Adressen den Fragebogen ausfüllen, wäre eine Erfassung der IP-Adressen ohne expliziten Hinweis ethisch nicht vertretbar.

- Informationen zu Browser und Betriebssystem sind für die Diagnose hilfreich. Ein Vorabtest des Fragebogens bei verschiedenen Browsern und Betriebssystemen ist zu aufwendig. Sind bestimmte Kombinationen auffallend häufig mit Abbrüchen verbunden, deutet dies auf ein technisches Problem hin. Insbesondere bei der Verwendung von Scriptsprachen und moderneren Webdesign-Techniken (CSS, Layers) muss man davon ausgehen, dass derartige Probleme auftreten können.

6.2. Kontrolle auf Eingabefehler und Konsistenz

Fehlende, fehlerhafte oder inkonsistente Angaben können bei jeder Form der Befragung vorkommen. Interaktive Befragungsformen wie das persönlich-mündliche Interview bieten durch die Kontrollinstanz Interview prinzipiell bessere Kontrollmöglichkeiten als nicht-interaktive. Die übermittelnde Person stellt andererseits selbst eine Fehlerquelle dar.

Ein Problem schriftlicher Befragungen ist die relative Leichtigkeit, mit der einzelne Fragen aufgrund des Fehlens sozialer Kontrolle übersprungen werden können. Hier lässt sich natürlich einwenden, dass keine Antwort häufig besser ist, als eine durch mehr oder weniger intensiven sozialen Druck erzwungene Antwort. Außerdem kann die Existenz des Interviewers auch die Auskunftswilligkeit des Befragten bei Fragen persönlicher Natur dämpfen, da die subjektiv empfundene Anonymität beeinträchtigt ist.

Fehlerhafte Antworten in einer Befragung können nicht nur durch mangelndes Verstehen oder fehlende Auskunftsbereitschaft seitens des Befragten entstehen, sondern auch durch triviale Schreibfehler bei der Erfassung durch den Interviewer bzw. den Befragten, oder im Prozess der Übertragung der Daten in den Computer. Bei Paper&Pencil-Umfragen können Fehler, die bei der Erfassung entstehen, nicht mehr korrigiert werden. Computergestützte Befragungsverfahren bieten den Vorteil, bestimmte Typen von Fehleingaben sofort zu bemerken und den Eingebenden zur Korrektur aufzufordern.

6.2.1. Fehlende und fehlerhafte Eingaben

Bei der Eingabe können zwei Typen von Fehlern auftreten:

- fehlende Werte: die Eingabe fehlt vollständig;
- „wilde Codes“: Die Eingaben liegen außerhalb des gültigen Bereichs.

Es ist sinnvoll, durch das Befragungs-Interface beide Fehlerarten abzufangen. „Wilde Codes“ können nur bei offenen Antworten – in der Regel numerischer Natur – auftreten, und sind in der Regel die Folge von Tippfehlern (wenn z.B. eine Taste doppelt betätigt wird, oder die gewünschte Taste verfehlt wird). Das Fehlen einer Angabe ist bei jeder Form der Eingabe möglich²⁵.

Die fehlende oder fehlerhafte Angabe wird vom Programm bemerkt, das den Befragten bittet, die Eingabe zu wiederholen. Das Erzwingen von Eingaben birgt allerdings Risiken – wenn ein Befragter eine Information – z.B. sein Einkommen oder seine parteipolitische Präferenz – nicht preisgeben möchte, wird er im Zweifelsfall die Befragung eher abbrechen – oder lügen.

Ein gutes Beispiel ist das Einkommen: Obwohl der Befragte davon ausgehen kann, dass seine Angaben anonym bleiben, besteht die Tendenz, das Einkommen nicht anzugeben. Wird der Befragte zur Angabe gezwungen, besteht die Gefahr dass er entweder falsche Angaben macht, oder aber die Befragung abbricht. Dies lässt sich durch die Angabe einer Residualkategorie vermeiden („keine Angabe“). Eine derartige Vorgehensweise empfiehlt sich allerdings nur bedingt, denn sie könnte dazu führen, dass bereits bei geringen Ressentiments gegenüber der Beantwortung die Residualkategorie gewählt wird. Lässt man die Überprüfung weg, dann verliert man natürlich auch die Antworten der Befragten, bei denen das Fehlen der Angabe auf Eingabefehler zurückzuführen ist.

Eine elegante, aber aufwendige Möglichkeit, ist eine Form der Abfrage, die auf evtl. Ressentiments eingeht: Wenn die Angabe zur kritischen Frage fehlt, wird der Befragte nicht einfach zur erneuten Eingabe aufgefordert, sondern erhält die Alternative, die Frage zu beantworten oder zur nächsten Angabe zu gehen.

Zu den Fragen, bei denen eine Eingabeüberprüfung auf jeden Fall notwendig ist, gehören die Filterfragen, denn für die korrekte Filterführung ist eine Informationsgrundlage notwendig.

²⁵ Außer bei einer Checkbox, bei der auch das Nicht-Ankreuzen eine gültige Angabe darstellt, und die aus diesem Grund natürlich eine Quelle für nicht zu kontrollierende Fehleingaben darstellt.

6.2.2. Konsistenzfehler

Konsistenzfehler entstehen, wenn Befragte an unterschiedlichen Stellen des Fragebogens widersprüchliche Angaben zu Sachverhalten machen, die logisch verknüpft sind.

Bsp.

- Nationalität != deutsch & bei der letzten Bundestagswahl gewählt.
- Wohndauer am Heimatort > Alter.

Konsistenzfehler können durch den Abruf vorheriger Angaben aus der Datenbank während der Eingabe festgestellt werden. Der Befragte sollte dann die Möglichkeit erhalten, eine der beiden fehlerhaften Angaben zu korrigieren.

Natürlich können Konsistenzfehler auch komplexere Formen annehmen. Z.B. könnte eine Aufsummierung der Ausgaben für Freizeitaktivitäten das frei verfügbare Einkommen überschreiten. Auch hier lässt sich eine programmgesteuerte Korrekturmöglichkeit anbieten. Eleganter ist es allerdings, wenn der Befragten dabei unterstützt wird, Konsistenzfehler zu vermeiden. Im angeführten Beispiel könnten z.B. die Angaben zu den Ausgaben vom Befragungsprogramm addiert werden und dem Befragten vor der Frage zum frei verfügbaren Einkommen als Erinnerungstütze präsentiert werden.

6.3. Das Problem des Befragungsabbruchs

Ein Befragter, dessen Zeitbudget von der Befragung stärker strapaziert wird, als er vorhergesehen hatte, wird eine Tendenz zum Abbruch der Befragung entwickeln. Bei einer Face-to-Face- oder Telefonbefragung wird diese Tendenz, wenn sie nicht zu stark wird, von der Höflichkeit gegenüber dem Interviewer bzw. dessen positive Motivierung zur weiteren Teilnahme ausgeglichen.

Auch das Auftreten von Fragen, die dem Befragten als zu persönlich, aufdringlich oder einfach nur „dumm“ erscheinen kann zum Abbruch führen.

Im Falle einer schriftlichen Befragung besteht die Chance, dass ein Fragebogen, dessen Bearbeitung aus Zeitgründen unterbrochen wird, weggelegt und später weiter ausgefüllt wird. Diese Möglichkeit fällt bei Online-Befragungen in der

Regel weg²⁶. Die Aufgabe, den Fragebogen möglichst kurz und wenig monoton zu halten, ist hier also noch wichtiger als bei anderen Befragungsformen.

Das Abbrecherproblem bei Online-Befragungen hat auch einen positiven Aspekt, nämlich bessere Möglichkeiten der Diagnose von Abbruchverhalten. Gerade durch das Fehlen sozialer Kontrolle und damit verbundenen niedrigen Hemmschwellen für den Abbruch lässt sich – möglichst in einem Pretest – präzise diagnostizieren, ob bei bestimmten Fragen oder einer bestimmten Dauer besonders häufig das Ausfüllen des Fragebogens eingestellt wird. Das schrittweise Abspeichern von Teilen der Befragung ermöglicht eine relativ detaillierte Analyse von Abbrecherdaten, die bei Pencil&Paper-Befragungen unmöglich ist.

Je häufiger die Fragen gespeichert werden (d.h., je weniger Fragen pro Seite vorhanden sind), desto genauer kann man analysieren, wann es zum Abbruch gekommen ist. Auf diese Weise kann man Fragen – oder Fragenkomplexe – isolieren, die zum Abbruch beigetragen haben können; durch das Speichern von Zusatzinformationen (Zeitpunkt des Speicherns) kann man anderen Gründen für einen Abbruch auf die Spur kommen²⁷.

²⁶ Programmtechnisch ist es möglich, eine Unterbrechungsmöglichkeit anzubieten, bei der der Befragte über ein Ticket oder einen speziellen Hyperlink die Befragung später wieder aufnehmen kann. Allerdings setzt die Wiederaufnahme einen gewissen Grad an Motivation voraus.

²⁷ Wenn die Zeiten für das Ausfüllen einer Seite vor dem Abbruch überdurchschnittlich zunehmen, kann man davon ausgehen, dass der Befragte abgelenkt wurde.

7. Einsatzfelder für Online-Befragungen

Die Tatsache, dass Computer mit Internet-Zugang in ihrer Verbreitung noch weit vom Telefon entfernt sind, macht die Online-Befragung für generelle Bevölkerungsumfragen ungeeignet. Die soziodemographischen Charakteristika der „Online-Population“, derjenigen, die ein Mindestmaß an regelmäßiger Nutzung des WWW aufweisen, sind nach wie vor weit von dem Bevölkerungsdurchschnitt entfernt.

7.1. Generelle Befragungen der Online-Population

Ist die „Online-Population“ eine Grundgesamtheit, deren Erforschung überhaupt interessant ist? Rückschlüsse auf Meinungen, Einstellungen und Verhalten der Bevölkerung lässt sie sicher nicht zu, aber die „virtuelle Welt“ des WWW ist inzwischen differenziert genug, um wissenschaftliche wie kommerzielle Interessen an ihren „Bewohnern“ entstehen zu lassen.

Für eine repräsentative Befragung der „Online-Population“ wird man aber trotzdem auf konventionelle Befragungstechniken, zumindest aber konventionelle Stichprobenverfahren, zurückgreifen müssen.

Welche Voraussetzungen müssen erfüllt sein, um eine Online-Befragung sinnvoll erscheinen zu lassen?

- Jedes Mitglied der Population hat Zugang zum Internet, alle Mitglieder weisen eine möglichst homogene Nutzung auf.
- Über die Population liegen genug Informationen vor, um eine Stichprobenziehung zu ermöglichen bzw. mittels einer Vollerhebung alle Mitglieder zu erreichen.
- Eine repräsentative Abbildung einer Population ist nicht Ziel der Befragung, sondern z.B.
 - die Evaluation eines Messinstruments;
 - das Sammeln von Informationen für die Exploration;
 - eine quasi-experimentelle Studie.

7.2. Mitarbeiterbefragungen

In einer Firma, in der alle Mitarbeiter über einen Zugang zum Inter- oder Intranet verfügen, ist ein Online-Fragebogen ein ökonomisches Mittel zur Durchführung einer Mitarbeiterbefragung.

- Alle Mitglieder der Population können entweder über E-Mail oder andere Kommunikationswege zuverlässig erreicht werden, so dass eine Vollerhebung oder einfache Zufallsstichprobe möglich ist.
- Die technische Ausstattung ist bekannt, so dass die Möglichkeiten der Fragebogenkonzeption daran angepasst werden können.
- Die prinzipielle Möglichkeit der De-Anonymisierung der Daten muss durch geeignete technische oder organisatorische Massnahmen ausgeschlossen werden.

7.3. Nutzung und Akzeptanz von Web-Auftritten und Online-Dienstleistungen

Eine naheliegende Einsatzmöglichkeit für Online-Befragungen sind Studien, die die Nutzung und Bewertung von Web-Auftritten bzw. die über sie angebotenen Dienstleistungen betreffen. Wenn die betroffene Nutzergruppe bei der Site registriert ist, ist eine Kontaktaufnahme per E-Mail (oder evtl. andere Kommunikationskanäle) möglich, bei „offenen“ Sites lassen sich Banner anbringen, die auf die Befragung hinweisen. Wie in 4.3 diskutiert, ist die erste Form der Kontaktaufnahme vorzuziehen. In jedem Fall ist prinzipiell sowohl eine Stichprobenziehung als auch eine Vollerhebung möglich. Im Fall der Bewerbung der Umfrage über ein Banner spielt allerdings der Zeitfaktor bei der Definition der Population eine Rolle – denn eigentlich erhält man hier nur einen Ausschnitt aus der Grundgesamtheit, der durch die Zeit, in der die Umfrage online zur Verfügung steht, begrenzt ist.

7.4. Fragebogenexperimente

Ein weiteres Anwendungsgebiet für Online-Befragungen sind Fragebogenexperimente. Dabei werden verschiedenen, zufällig ausgewählten Gruppen unterschiedliche Fragebogenversionen präsentiert. Auf diese Art und Weise kann man den Einfluss von Frageformulierungen oder –anordnungen analysieren. Die Repräsentativität der Gruppen spielt dabei eine nachgeordnete Rolle –

wichtiger ist, dass der Unterschied zwischen den Gruppen möglichst klein, vor allem aber nur durch den Zufall bestimmt ist. Die Randomisierung von Gruppen, die bei Experimenten von großer Bedeutung ist, lässt sich programmtechnisch sehr einfach realisieren. Die im Rahmen von Fragebogenexperimenten gewonnenen Erkenntnisse lassen sich zur Verbesserung der Qualität von Befragungen nutzen.