

## Zur Verallgemeinerbarkeit empirischer Befunde bei nichtrepräsentativen Stichproben

Ein Problem sozialwissenschaftlicher Begleitung von Modellversuchen und Pilotprojekten, illustriert am Bildschirmtext-Feldversuch Düsseldorf/Neuss

Helmut Kromrey

Wie in fast allen empirischen Forschungsansätzen, so besteht auch bei der sozialwissenschaftlichen Begleitung von Modellversuchen und Pilotprojekten ein explizites Ziel – neben anderen – darin, die Befunde über den eigentlichen Analysegegenstand hinaus zu verallgemeinern. Als wesentliche Voraussetzung für jede Verallgemeinerung gilt vielen Forschern die Realisierung einer repräsentativen Stichprobe von Untersuchungsfällen. Im statistischen Sinne "repräsentative", gar nach dem Zufallsprinzip zustande gekommene Auswahlen von Teilnehmern an Pilotprojekten und Feldversuchen sind jedoch in aller Regel grundsätzlich nicht erreichbar. An solchem Mangel an Stichproben-Repräsentativität leidet im übrigen nicht nur der hier behandelte spezielle Untersuchungstyp, sondern angesichts abnehmender Antwortbereitschaft von Befragten jede Form von Umfrageforschung.

Der vorliegende Beitrag diskutiert Wege, die aus diesem Verallgemeinerungs-Dilemma herauszuführen können, und verdeutlicht diese am Beispiel der sozialwissenschaftlichen Begleitung zu den Bildschirmtext-Feldversuchen.

As in almost all empirical research approaches, one explicit objective inter alia of social scientific backup research on model-researching and pilot projects is to generalise findings beyond the scope of the specific object of analysis. For many researchers an essential requirement for any generalisation of participants in pilot projects and field research which have in a statistical sense been conducted "representatively", or even according to the principle of randomness are as a rule not generally attainable. The special research category dealt with in this article is not the only one to suffer from this lack of sample representativeness. In view of the declining response willingness of interviewees every form of opinion survey research is also adversely affected.

The following article discusses possible means of overcoming this generalisation dilemma, illustrating this via exemplary reference to backup social research on Bildschirmtext (Btx = interactive videotex) field trials.

### Vorbemerkungen zum Begriff der Repräsentativität

Empirische sozialwissenschaftliche Untersuchungen – ganz gleich welcher wissenschaftstheoretischen Tradition sich die Forscher verpflichtet fühlen – werden üblicherweise nicht rein deskriptiv unternommen, sondern mit dem Ziel, die Befunde in irgendeiner Weise über die tatsächlich untersuchten Sachverhalte hinaus zu generalisieren. Das gilt auch für die sozialwissenschaftliche Begleitung von Erprobungen neuer technischer Systeme, neuer Organisationsformen, neuer Curricula u. ä. in Modellversuchen oder Pilotprojekten.

Bei der Bewertung der Güte derartiger Untersuchungen spielt der Begriff "Repräsentativität" eine herausragende Rolle. Nicht-repräsentative Befunde – so ist häufig die einfache Gleichung – sind nicht verallgemeinerbar.

Leider ist jedoch das Konzept "Repräsentativität" in solchen Zusammenhängen nicht nur sehr beliebt, sondern auch ausgesprochen vieldeutig in seiner Verwendung – diesmal durchaus abhängig von der wissenschaftstheoretischen Tradition des Forschers, aber auch von seiner Bereitschaft, Argumente differenziert zu explizieren.

### Hauptverwendungen von "Repräsentativität"

In mehreren Beiträgen in der International Statistical Review berichten Kruskal und Mosteller (1) über die Verwendung dieses Begriffs in nichtwissenschaftlichen, in nichtstatistischen wissenschaftlichen sowie in statistischen Texten. Sie arbeiten nach umfangreicher Durchsicht von Materialien neun Hauptverwendungen heraus:

- Repräsentativität als ein nicht weiter präzisiertes Attribut, das der Autor den eigenen Daten und/oder Schlussfolgerungen zuschreibt, um den Leser in positiver Weise zu beeinflussen, um seinen eigenen Ausführungen das Siegel der Glaubwürdigkeit zu verleihen;
- Repräsentativität als *relativ* unbestimmt gelassener Terminus, der lediglich mit dem Hinweis auf ein nicht näher beschriebenes Auswahlverfahren begründet wird, den Leser aber weitgehend im unklaren darüber läßt, wie die angeblich repräsentative Auswahl zustande gekommen ist;
- Repräsentativität als Abwesenheit verzerrender, einseitig selektierender Faktoren im Auswahl- und Datenerhebungsprozeß; verzerrende oder einseitig selektiv wirkende Faktoren wären z. B. die Selbstrekrutierung von Teilnehmern an Experimenten bzw. die ungleiche Erreichbarkeit oder Antwortbereitschaft von Bevölkerungsstellen bei Interviews;
- Repräsentativität als die Eigenschaft einer Stichprobe, ein verkleinertes Abbild einer angebbaren Grundgesamtheit zu sein, ein miniaturisiertes Spiegelbild hinsichtlich aller für die Untersuchungszwecke bedeutsamen Merkmale und Merkmalskombinationen;
- Repräsentativität als Tendenz, als Ergebnis, das mit angebbarer Wahrscheinlichkeit bei einem ganz bestimmten Stichprobenverfahren – nämlich der Zufallsauswahl – eintritt.

In statistisch orientierten Argumentationen sind die drei letztgenannten Aspekte vor herrschend und werden üblicherweise im Zusammenhang gesehen: Wenn mit dem Blick auf eine präzise abgrenzbare Grundgesamtheit mit endlicher Zahl von empirisch definierbaren Elementen ein Plan zur Durchführung einer *kontrollierten Zufallsauswahl* erstellbar ist, und wenn bei der Durchführung der Stichprobenziehung sowie bei der Datenerhebung *verzerrende Einflüsse ausgeschaltet* werden können, dann ist das Ergebnis mit angebbarer Wahrscheinlichkeit ein *verkleinertes Abbild* dieser Grundgesamtheit. Das Sample ist dann mit angebbarer Wahrscheinlichkeit repräsentativ hinsichtlich sämtlicher Merkmale und Merkmalskombinationen – mit der (gewünschten) Konsequenz, daß aus den Stichprobenstatistiken auch auf *unbekannte* Parameter der Grundgesamtheit geschlossen werden kann.

Dieser letzte Aspekt wird in einer weiteren Verwendung des Begriffs in den Vordergrund gestellt, nämlich:

- Repräsentativität als Eigenschaft einer Stichprobe, die gute statistische Schätzungen für möglichst viele Parameter erlaubt. Dieses Kriterium ist beispielsweise ausdrücklich in der Definition zum Allbus 1980 (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) genannt: Wenn es "für möglichst viele ... durch die Untersuchungsziele definierte Partitionen ein *gemeinsames* Schätzverfahren gibt, mit dem

(1) Kruskal, William F., Mosteller: Representative Sampling. I: Non-Scientific Literature; II: Scientific Literature; III: Current Statistical Literature; IV: The History of the Concept in Statistics. In: International Statistical Review, Vol. 47, 1979, S. 13–27, 111–127, 245–265; Vol. 48, 1980, S. 169–194

zumindest die größeren Anteile (der Partitionen, H. K.) hinreichend genau geschätzt werden können, dann sei der aus dem Stichprobenplan resultierende Querschnitt *repräsentativ* genannt" (2). Der Grad solcher Repräsentativität ist kontrollierbar durch Vergleich bekannter Merkmalsverteilungen in der Grundgesamtheit mit den vorfindbaren Stichprobenverteilungen.

Auf der gleichen Ebene, jedoch mit geringerem Anspruch, bewegt sich eine siebte Begriffsverwendung:

– Repräsentativität als Attribut einer Stichprobe, die für *bestimmte, eingeschränkte Zwecke* als "gut genug" gilt. Wenn es z. B. um die Prüfung einer These geht, daß eine bestimmte Erscheinung in der Grundgesamtheit gar nicht oder fast gar nicht vorkommen dürfte, dann ist zur Widerlegung dieser These eine Stichprobe hinreichend, die diese Erscheinung mit einiger Häufigkeit nachweist.

Ganz von der statistischen Argumentationslinie entfernen sich zwei weitere Verwendungen des Konzepts, die sich an der Variation und Heterogenität der repräsentierten *Merkmale* bzw. an der Sammlung von *typischen Fällen* orientieren.

– Nach der ersten dieser Vorstellungen bedeutet Repräsentativität, daß die Stichprobe eine möglichst breite Vielfalt unterschiedlicher Elemente aus der Grundgesamtheit enthält, so daß sich die vollständige Heterogenität von Merkmalsausprägungen und Merkmalskombinationen auch in der Stichprobe wiederfindet. Im Unterschied zum Modell des verkleinerten Abbilds der Grundgesamtheit wird jedoch *nicht* eine *maßstabgerechte* Repräsentation der Häufigkeit von Merkmalsausprägungen gefordert.

– Die Gleichsetzung von "repräsentativ" mit "typisch" schließlich bedeutet, daß sich Repräsentativität nicht erst an den globalen Eigenschaften der Stichprobe äußert, sondern sich an jedem einzelnen Fall zu erweisen hat. Jedes einzelne Element der Stichprobe soll ein typischer Vertreter für eine Gruppe ähnlicher Elemente in der Grundgesamtheit sein. "Typisch" kann dabei definiert sein durch die Nähe zu einem Idealtypus oder durch die Häufigkeit des Auftretens oder durch einen Durchschnittswert oder durch eine spezifische Kombination von Merkmalen.

Auch die beiden zuletzt genannten Kriterien lassen sich in den Anforderungen an einen Stichprobenplan kombinieren: Man definiere für eine zu untersuchende Fragestellung ein möglichst breites Spektrum relevanter Merkmalskombinationen (Kriterium: Heterogenität) und wähle für jede dieser Merkmalskombinationen mindestens einen typischen Fall aus. Fügt man als drittes Kriterium noch die Forderung nach maßstabgetreuer Repräsentation der Häufigkeit des Auftretens themenrelevanter Merkmalskombinationen in der Grundgesamtheit hinzu, so gelangt man zum oben bereits genannten Repräsentativitätsmodell: Stichprobe = verkleinertes Abbild der Grundgesamtheit; diesmal allerdings nicht realisiert durch ein Zufallsample, sondern durch die gezielte Auswahl typischer Fälle.

Schon diese Übersicht über hauptsächlichliche Verwendungen des Repräsentativitätsbegriffs dürfte zeigen, daß eine einfache und zugleich allgemeingültige Antwort auf die Frage der Generalisierung empirischer Befunde aus sogenannten nicht-repräsentativen Stichproben nicht zu erwarten ist. Der Sprachgebrauch wird zudem noch weiter verkompliziert, wenn – wie etwa bei Holzkamp (3) – zwischen maximaler, optimaler

(2) Kirschner, Hans-Peter, ALLBUS 1980: Stichprobenplan und Gewichtung. In: Mayer, K.-U./P. Schmidt (Hg.): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, Frankfurt/M., New York 1984, S. 136  
(3) Holzkamp, Klaus: Theorie und Experiment in der Psychologie, Berlin 1964, S. 31ff.

und faktischer Repräsentanz unterschieden wird. Und der gemeinte Sachverhalt wird noch weiter verwirrt, wenn – vom gleichen Autor – der Begriff nicht nur auf die (in welchem Sinne auch immer) angemessene Repräsentation der Untersuchungsobjekte bezogen, sondern erheblich ausgeweitet und zusätzlich untergliedert wird in Subjekt-, Umgebungs-, Handlungs- und Erlebnisrepräsentanz sowie der Stichprobenrepräsentanz als einer Sonderform, die sich mit den vorhergehenden Konzepten teilweise überschneidet (4).

#### Definition von "Repräsentativität" in diesem Beitrag

Trotz – oder gerade wegen – seiner insgesamt widersprüchlichen Handhabung läßt sich dieses methodologische Konzept auch im hier vorliegenden Diskussionsbeitrag nicht vermeiden. Daher sei klargestellt: Wenn im folgenden von "Repräsentativität" die Rede ist, dann in der Bedeutung, wie sie sich in der Methodendiskussion traditioneller empirischer Sozialforschung weit überwiegend eingebürgert hat. Unter "Repräsentativität" ist danach zu verstehen – so eine zusammenfassende Definition bei Böltken –, "daß in einer Auswahl alle für die Grundgesamtheit ... charakteristischen Merkmale und Merkmalskombinationen getreu ihrer relativen Häufigkeit vertreten sein müssen und somit die Auswahl ein verkleinertes Abbild der Grundgesamtheit selbst für solche Merkmale ist, von deren Vorhandensein wir (noch) gar nichts wissen" (5).

Als in diesem Sinne "nicht-repräsentativ" haben dann alle Stichproben zu gelten, – die nicht nach dem Zufallsverfahren gezogen wurden (wegen der Forderung, auch vorab unbekannte Merkmale zu repräsentieren), und/oder  
– bei deren Realisierung verzerrende, einseitig selektierende Einflüsse wirksam wurden (weil damit der Charakter des Zufalls-Samples zerstört würde), wodurch dann auch  
– keine statistisch abgesicherten Schätzungen von Grundgesamtheits-Parametern möglich wären.

#### Forschungsfeld: sozialwissenschaftliche Begleituntersuchungen

Ich möchte mich nun hier nicht im einzelnen damit auseinandersetzen ob es a) überhaupt *denkbar* ist, für sozialwissenschaftliche Untersuchungen in nennenswertem Umfang Stichprobenpläne zu entwickeln, die diesen Vorstellungen in vollem Umfang gerecht werden, und falls ja, ob b) die Durchführung der Stichprobeneziehung überhaupt in einer Weise *möglich* ist, den Stichprobenplan zumindest mit hinreichender Annäherung empirisch zu realisieren.

Ich befaße mich mit einem Forschungsfeld, in dem beides von vornherein ziemlich *ausgeschlossen* scheint: dem Typ empirischer Sozialforschung, der mit Schlagworten wie Erfolgskontrolle, Evaluation, Akzeptanzforschung, Folgenabschätzung, Begleitforschung umrissen wird. Das sind etwa Forschungen im Rahmen des Programms "Humanisierung des Arbeitslebens" oder (aktuell) des NRW-Programms "Mensch und Technik"; Begleitforschungen zum "Experimentellen Wohnungs- und Städtebau" des Bundesbauministeriums; Folgenabschätzungen der Einführung neuer Produktions-techniken oder Begleituntersuchungen bei der Erprobung neuer IuK-Techniken; aber auch z. B. Evaluation von Gesamtschulversuchen. Die Liste ließe sich fortsetzen.

(4) aaO, S. 97ff.

(5) Böltken, Ferdinand: Auswahlverfahren, Stuttgart 1976, S. 128. – Diese Definition bezieht sich auf eine "einfache", d. h. nicht geschichtliche Stichprobe. Im Falle geschichtlicher Sample müssen anhand bekannter Auswahlsätze die relativen Häufigkeiten rekonstruierbar sein.

In der Regel wird in solchen Fällen der Forscher vor ein *Verallgemeinerungs-Dilemma* gestellt. Fast immer fordert nämlich der Auftraggeber sozialwissenschaftlicher Begleitstudien auch die Generalisierbarkeit der empirischen Befunde, also die Loslösung der Ergebnisse vom Kontext ihrer Entstehung, von den möglicherweise sehr spezifischen Bedingungen und Besonderheiten des jeweiligen Untersuchungsgegenstands. Und fast immer findet sich dann unter den Kritikpunkten zu den veröffentlichten Forschungsberichten an prominenter Stelle auch der Vorwurf mangelnder Repräsentativität; gefolgt von der ebenso regelmäßig daraus abgeleiteten Schlußfolgerung, daß damit die empirischen Befunde allenfalls als eine deskriptive Diagnose des jeweiligen eng begrenzten Untersuchungsgegenstands gelten könnten, aber – bitte sehr – nicht darüber hinausgehend verallgemeinert werden dürfen.

Der Mangel an Repräsentativität – im oben definierten statistischen Sinne – ist natürlich leicht zu belegen. Das fängt schon bei der Konstruktion des Stichprobenplans an:

– Eine zentrale Voraussetzung für die *Entwicklung von Stichprobenplänen* ist, daß die Grundgesamtheit der Fälle, auf die die Befunde verallgemeinert werden sollen, genau definierbar und empirisch angebar sein muß. Dies ist aber weder für die Anwendungsfälle des im Pilotprojekt erprobten Systems möglich noch für die von diesem System Betroffenen. Die Grundgesamtheit existiert zum Zeitpunkt der Untersuchung noch gar nicht; ja, es sind zu Beginn der Untersuchung noch nicht einmal präzise Angaben über potentielle künftige Anwendungsfälle sowie potentielle künftige Nutzer oder in anderer Weise Betroffene möglich. Darauf soll nämlich die durchzuführende Studie (unter anderem) empirisch begründete Hinweise liefern. Damit ist natürlich von vornherein auch eine Zufallsauswahl von zu untersuchenden Anwendungsfällen oder von Betroffenen ausgeschlossen.

– Aber selbst wenn die Gesamtheit der Anwendungsfälle vorab angebar wäre (etwa bei Plänen zur flächendeckenden Einführung eines neuen Fernmeldedienstes oder eines neuen Schulsystems bei gleichzeitiger Einstellung der bisherigen Angebote) und selbst wenn daraus die Gesamtheit der potentiellen künftigen Betroffenen ableitbar wäre, stünden der *Entwicklung eines auf statistische Repräsentativität angelegten Forschungsdesigns* unüberwindliche Hindernisse im Wege: Die Erprobungen könnten (wegen der damit verbundenen Investitionen) nicht flächendeckend, sondern nur an sehr wenigen, unter pragmatischen Gesichtspunkten zu bestimmenden Orten durchgeführt werden; und sie könnten dort (wegen des Erprobungscharakters) lediglich parallel zu den bisherigen Angeboten organisiert werden, könnten diese also (noch) nicht ersetzen. Damit ist es – wiederum – nicht möglich, eine Zufallsstichprobe von zwangsläufig durch diese Änderung Betroffenen zu ziehen; vielmehr müssen Interessenten gefunden werden, die sich freiwillig an der Erprobung beteiligen. An die Stelle einer forschungsmethodisch begründeten und kontrollierten Auswahl wird somit eine erprobungspragmatisch begründete Auswahl mit Selbstrekrutierung der Teilnehmer zu treten haben.

– Konsequenz: Ein auf Repräsentativität im herkömmlichen statistischen Sinne angelegtes Design mit Zufallsauswahl ist *grundsätzlich* nicht möglich; allenfalls kann man sich um Annäherungen bemühen.

### Beispiel Bildschirmtext

Ich möchte die Konsequenzen solcher Rahmenbedingungen, denen sich die Forschung wohl oder übel zu unterwerfen hat, am Beispiel der sozialwissenschaftlichen Begleitung zum Feldversuch Bildschirmtext in Düsseldorf/Neuss (getragen und finanziert von der

Deutschen Bundespost und dem Land Nordrhein-Westfalen) illustrieren. Ich werde auch im weiteren vor allem auf dieses Beispiel zurückgreifen, weil mir hierzu aus der eigenen intensiven Beteiligung die differenziertesten Daten zur Verfügung stehen.

Das genannte Verallgemeinerungs-Dilemma bestand auch in diesem Fall. Im nordrhein-westfälischen Bildschirmtext-Versuchsgesetz heißt es in § 3, Abs. 1:

„Die Landesregierung hat sicherzustellen, daß der Feldversuch wissenschaftlich begleitet und ausgewertet wird. Neben den Auswirkungen im Medienbereich sind insbesondere auch die sozialen, kulturellen und wirtschaftlichen Folgen zu untersuchen.“

Es sollten also die Auswirkungen des zu erprobenden neuen Systems untersucht werden, und zwar nicht beschränkt auf die am Pilotprojekt *teilnehmende* Gruppe von Personen und Haushalten; vielmehr wird generell formuliert: soziale, kulturelle und wirtschaftliche Folgen. Es fehlt an jeder Einschränkung der Reichweite auf den Feldversuch. Das ist verständlich, denn die empirischen Befunde aus dem Feldversuch sollten ja – so § 3, Abs. 3 des Btx-Versuchsgesetzes NW – die Grundlage der Beschlußfassung des Landtags über die endgültige Einführung von Bildschirmtext und über die Bedingungen sein, unter denen das System dann betrieben werden darf. Folgerichtig war daher zunächst auch geplant, eine für den Raum Düsseldorf/Neuss möglichst bevölkerungsrepräsentative Stichprobe von Teilnehmern zusammenzustellen. Teilnehmerkreis und Teilnahmebedingungen stellten sich in der vorherigen Pilotprojekt-Planung wie folgt dar:

*Teilnehmer* am Feldversuch Bildschirmtext Düsseldorf/Neuss:

- Anbieter von Informationen und anderen Diensten (Genehmigung durch den Regierungspräsidenten Düsseldorf erforderlich);
- private Haushalte im Feldversuchsgebiet (Sollzahl: 2000, bevölkerungsrepräsentativ für den Versuchsraum); geschichtete Zufallsauswahl, aus dem Kreis der Interessenten bevölkerungsrepräsentativ ausgewählt;
- 1000 weitere Teilnehmer, die von den Informationsanbietern benannt werden.

*Teilnahmebedingungen* für die privaten Haushalte (= Zielgruppe der wissenschaftlichen Begleituntersuchung):

- Rücksendung einer Antwortkarte an die Bundespost in Reaktion auf Postwurf- und Anzeigenwerbung für die Pilotprojekt-Teilnahme (= Bekundung von Interesse an der Teilnahme);
- Auswahl der Teilnahme-Berechtigten aus den Rückmeldern nach dem Prinzip der geschichteten Zufallsauswahl (Bezugspopulation: Haushalte im Versuchsraum Düsseldorf/Neuss);
- Zusendung eines „Btx-Kaufgutscheins“ durch die Bundespost an die ausgewählten Haushalte (= Berechtigung zum Erwerb eines Btx-tauglichen Fernsehers zum Preis eines marktüblichen TV-Geräts der oberen Preisklasse).

Faktisch allerdings war eine Zufallsrepräsentativität (für großstädtische Haushalte) auch ansatzweise nicht erreichbar (6). Selbst bei Ausschöpfung *aller* privaten Haushalte, die unter den oben angeführten Bedingungen teilnahmebereit waren, würde mit

(6) Zum realisierten Design vgl. Jansen, Dorobica/H. Kromrey/H. Treinen: Bildschirmtext – Fernmeldedienst oder neues Medium. In: Rundfunk und Fernsehen, 1982/4, S. 447–465; Kromrey, Helmut: Bildschirmtext – ein neues Kommunikations- und Informationssystem. In: Mass communication, 1984/1, S. 12–24; Kromrey, Helmut/H. Treinen: Sozialwissenschaftliche Begleituntersuchung zum Feldversuch Bildschirmtext Düsseldorf/Neuss. In: Angewandte Sozialforschung, 1983/1–2, S. 141–144.

rund 1600 Versuchshaushalten nicht einmal die geplante Sollzahl erreicht. Das Teilnehmerfeld setzte sich also nach dem Prinzip der Selbstselektion zusammen: Jeder, der wollte, konnte mitmachen.

Die einzelnen Schritte der *Teilnehmerwerbung* verdeutlichen diesen Fehlschlag des ursprünglichen Plans:

- Interessentenwerbung und Auswahl:  
Auf zwei Postwurfsendungen sowie auf zusätzliche Anzeigenwerbung meldeten sich anstelle der von der Bundespost erwarteten mehreren Zehntausend nur rund 2800 Teilnahme-Interessenten. Hierbei waren an- und ungelernete Arbeiter, Facharbeiter und einfache Angestellte gegenüber dem Bevölkerungsquerschnitt im Versuchsraum unterrepräsentiert. Zum Ausgleich dieser Verzerrung in der Interessentenbasis wurden *sämtlichen* Haushalten von Arbeitern und einfachen Angestellten, die eine Rückantwortkarte eingeschickt hatten, Btx-Kaufgutscheine zugeteilt (= Auswahl-satz 100 %). Aus dem im Interessentenkreis überrepräsentierten Gruppen wurde eine Zufallsauswahl getroffen, allerdings mit sehr hohem Auswahl-satz, um insgesamt an die Sollzahl 2000 heranzukommen.

- Ausfälle:

Allerdings lösten nicht alle ursprünglich an der Teilnahme Interessierten ihren Kaufgutschein ein, d. h. sie setzten ihr bekundetes Teilnahme-Interesse nicht in den Kauf eines neuen, Btx-tauglichen Fernsehers um. Die Zahl der sogenannten "Abspringer" war relativ groß. Unter den tatsächlichen Feldversuchs-Teilnehmern (den Btx-Geräte-käufern) waren Arbeiter und einfache Angestellte noch stärker unterrepräsentiert als schon bei den Teilnahme-Interessenten.

- Reine Selbstselektion:

Daher wurde schließlich jede Absicht eines bevölkerungsrepräsentativ geschichteten Samples gänzlich aufgegeben und auch dem verbleibenden Teil der Interessenten der Kaufgutschein zugeschickt.

*Konsequenz* dieses Verlaufs der Teilnehmerwerbung: Deutlich überrepräsentiert waren im Teilnehmerfeld im Vergleich zur Bevölkerung des Versuchsraums Personen und Haushalte mit den Merkmalen:

- weiterführende Schulbildung,
- qualifizierte Berufsposition,
- hohes Einkommen, mehrere Einkommensbezieher im Haushalt,
- überdurchschnittliche Ausstattung mit informationstechnischen Geräten,
- breites Nutzungsspektrum an Printmedien, insbesondere Fachzeitschriften,
- überdurchschnittliche Bedarfsbreite und Nachfrageintensität an unterschiedlichsten Informationsarten.

*Erstes Fazit:* Von Bevölkerungsrepräsentativität keine Spur!

Dementsprechend ließ auch die Kritik nicht lange auf sich warten; man monierte, daß schon aus diesem Grunde die Ergebnisse der Begleitforschung ziemlich wertlos seien. Solche Kritik kommt nun nicht etwa nur von forschungsmethodischen Laien oder von wissenschaftstheoretischen Puristen, sondern auch von pragmatisch-anwendungsorientierten Sozialwissenschaftlern. So bilanzieren z. B. Hellstern/Wollmann, weder die privaten und gewerblichen Teilnehmer in beiden Versuchsgebieten (Düsseldorf und Berlin) noch die Informationsanbieter seien für künftige Nutzer und spätere Anbieter repräsentativ. Und weiter: "Auch im Hinblick auf die Gesamtbevölkerung weichen die

durch Selbstrekrutierung entstandenen Teilnehmergruppen erheblich vom Bevölkerungsdurchschnitt ab... Ergebnisse lassen sich daher kaum verallgemeinern." (7)

Meine These, die ich dagegensetzen möchte, ist: Die Nicht-Repräsentativität von Stichproben (Nicht-Repräsentativität im zu Beginn definierten Sinn) ist für sich genommen noch *kein* hinreichendes Argument gegen die Möglichkeit, die Befunde zu verallgemeinern.

Ausgeschlossen ist zunächst lediglich die sozusagen "automatische" Verallgemeinerung mit Hilfe wahrheitsstatistischer Verfahren.

### Verallgemeinerungslogik bei Experimenten

Wie vorzeitig die häufig unreflektierte Gleichsetzung einer solchen Nicht-Repräsentativität von Untersuchungseinheiten mit der (behaupteten) Unmöglichkeit von Verallgemeinerungen ist, zeigt schon ein Vergleich mit der Methodologie des sozialwissenschaftlichen Experiments. Gerade das Experiment gilt ja als Untersuchungsdesign, das für *Kausalanalysen* in idealer Weise geeignet sein soll; und genau darum geht es bei der Untersuchung der Frage, welche Auswirkungen bestimmte Maßnahmen haben und wie Ursachen und Wirkungen in Richtung und Stärke zusammenhängen.

Die folgende Übersicht zeigt einige der wichtigsten Experimental-Designs.

### Design-Strukturen bei Experimenten (8)

four-group design (Solomon)	pretest-posttest controlgroup design	G1 R	M1 × M2	(non) equivalent control group design
		G2 R	M1 - M2	
	posttest-only controlgroup design	G3 R	- M	static group-comparison
		G4 R	× M	
		one-shot case study		

mit Randomisierung

ohne Randomisierung

Erläuterung: G = Gruppe; R = Randomisierung; M = Messung; × = "treatment"

Das Design eines "echten" Experiments zeichnet sich dadurch aus, daß es mindestens die folgenden Merkmale aufweist:

- Es existiert eine Experimentalgruppe G1, die dem experimentellen Stimulus X, dem "treatment" (der auf ihre Auswirkungen zu untersuchenden Maßnahme) ausgesetzt wird.
- Es existiert eine in allen wesentlichen Merkmalen äquivalente Kontrollgruppe G2, die dem experimentellen Stimulus *nicht* ausgesetzt wird, die von der Maßnahme "verschont" bleibt.

(7) Hellstern, Gerd-M./H. Wollmann: Bilanz - Reformexperimente, wissenschaftliche Begleitung und politische Realität. In: dies.: Experimentelle Politik - Reformstrohfeuer oder Lernstrategie. Opladen 1983, S. 49f.

(8) nach Frey, Siegfried/H.-G. Frenz: Experiment und Quasi-Experiment im Feld. In: *Paed. / Soc. / Inc. (Hrsg.)* Feldforschung, Bern 1982, S. 250

- In beiden Gruppen werden vor dem Zeitpunkt des treatments und ausreichende Zeit danach die Ausprägungen der abhängigen Variablen (die Merkmale, bei denen man Auswirkungen durch die Maßnahme erwartet) gemessen (M1 und M2).  
- Stimmen vor dem treatment in der Experimental- und in der Kontrollgruppe die Verteilungen der abhängigen Variablen überein (und das sollten sie bei äquivalenten Kontrollgruppen), und sind nach dem treatment Unterschiede zwischen den Gruppen feststellbar, dann werden diese Unterschiede als Effekte des treatments (als Auswirkungen der Maßnahmen) interpretiert.

Dieses Design kann noch um zwei weitere Gruppen (eine Experimental- und eine Kontrollgruppe, G3 und G4) erweitert werden, in denen man auf die Messung vor dem treatment verzichtet. Dadurch wird kontrolliert, ob nicht allein durch die Vorhermessung schon Veränderungen in Gang gesetzt wurden (Versuchskaninchen-Effekt). Für Untersuchungsgegenstände, bei denen man in Bevölkerungs-Teilgruppen jeweils unterschiedliche Auswirkungen der gleichen Maßnahmen für möglich hält (z. B. alte Leute gegenüber Jugendlichen, Frauen gegenüber Männern, Familien mit Kleinkindern gegenüber alten Ehepaaren usw.), kann außerdem das Design auf eine größere Zahl von Experimental- und zugeordneten Kontrollgruppen ausgeweitet werden (für jede relevante Bevölkerungs-Teilgruppe ein komplettes Experimentaldesign).

Als Prototyp des sozialwissenschaftlichen Experiments ist sicher das Labor-Experiment anzusehen. Hierfür gilt, daß - im günstigsten Fall - die Auswirkungen möglichst aller Randbedingungen mit Ausnahme des experimentellen Stimulus bekannt sein sollten, so daß die Äquivalenz von Versuchs- und Kontrollgruppe auf der Basis empirisch bestätigter theoretischer Kenntnisse herstellbar ist. Die Zusammensetzung der Gruppen kann also in einem solchen Fall gezielt vorgenommen werden, die möglichen Einflußgrößen sind gezielt kontrollierbar. Um präzise verallgemeinerbare Ergebnisse zu erzielen, würde man auf eine Zufallsauswahl der Untersuchungspersonen gerade bewußt verzichten.

Diese anspruchsvolle Voraussetzung - daß man schon hinreichende empirisch bestätigte Kenntnisse hat, um kontrollierte Experimente durchzuführen - ist allerdings normalerweise nicht gegeben. Daher machen sich auch sogenannte echte sozialwissenschaftliche Experimente den Vorteil des kontrollierten Zufallsprinzips zunutze, der darin besteht, auch (noch) unbekannt Merkmale und Faktoren mit angegebener Wahrscheinlichkeit in einer nicht einseitig verzerrenden Weise zu repräsentieren. Allerdings wird das Zufallsprinzip gerade nicht auf die Auswahl der Experimentalt Teilnehmer angewendet - hier ist man in der Regel auf Freiwillige angewiesen; man hat es also selbst im echten Experimentaldesign (nicht nur in Quasi-Experimenten) in der Regel mit einer systematisch verzerrten Teilnehmerstichprobe zu tun. Das Zufallsprinzip wird eine Stufe tiefer eingesetzt: Sobald eine genügende Anzahl von Experimentalt Teilnehmern gefunden ist, werden diese nach den Merkmalen geschichtet, die für den Ausgang des Experiments als bedeutsam gelten - vielleicht Alter, Geschlecht, Bildung; danach entscheidet ein Zufallsverfahren, welche Personen aus jeder Schicht den Experimentalgruppen und welche den Kontrollgruppen zugewiesen werden.

Auf diese Weise erreicht man zweierlei: zum einen die Bildung von äquivalenten Experimental- und Kontrollgruppen hinsichtlich der Schichtungsmerkmale, zum anderen durch das zufällige Zuweisen zu Experimental- oder Kontrollgruppe (durch das sogenannte Randomisieren) die Ausschaltung der Gefahr systematischer Ergebnisverzerrungen durch andere, dem Forscher vorab nicht bekannte Faktoren. Das Zufallsverfahren dient somit nicht dazu, eine möglichst weitgehende Identität zwischen Grundgesamtheit und verkleinertem Abbild (Stichprobe) zu erreichen, sondern die

möglichst weitgehende Identität zwischen Experimental- und Kontrollgruppen zu sichern.

Entscheidend für die Verallgemeinerbarkeit von Resultaten kann dementsprechend auch nicht die Repräsentativität der Teilnehmerschaft für eine Grundgesamtheit von Personen sein, sondern die Frage, inwieweit die durch das Experiment konstruierte Realität mit der externen Realität, der Realität außerhalb des Forschungszusammenhangs übereinstimmt, also "externe Validität" aufweist. Eine gängige Definition lautet etwa: "Extern valide ist eine Untersuchung dann, wenn es möglich ist, ihre Ergebnisse auf andere Situationen, Personen, Verhaltensweisen und Treatments zu generalisieren..." (9).

H. Kreuz (10) hat im übrigen diesen Gedankengang mit dem Begriff des "Realitätsgehalts" von Forschungsdaten auf jede Form der Datenerhebung ausgeweitet. Er fordert als Kriterium für jede Verallgemeinerbarkeit eine hinreichende Ähnlichkeit der Forschungskontaktsituation mit einer - wie er sie nennt - "Realkontaktsituation", sofern die Daten nicht von vornherein in Realkontaktsituationen erhoben werden können.

Für die Verallgemeinerbarkeit empirischer Befunde ist aus solcher Perspektive die Repräsentativität einer Stichprobe von Fällen nur eine mögliche Voraussetzung unter mehreren, die zudem nicht einmal generell, sondern allenfalls in Spezialfällen erfüllt sein muß. Für das Experiment formuliert dies Holzkamp ganz explizit: "Die Frage danach, wie weit eine bestimmte Art von 'experimenteller Realität' für die entsprechende Art 'theoretischer Realität' 'repräsentativ' ist, läßt sich natürlich nur soweit sinnvoll stellen, als über diese Realität im 'theoretischen Satz' überhaupt etwas ausgesagt ist. So müssen in einem theoretischen Satz etwa überhaupt irgendwelche Klassifizierungen von Menschen enthalten sein, wenn das Problem der 'Subjektrepräsentanz' aktuell werden soll..." (11). Ähnliche Vorstellungen entwickelt Brunswik (12) im Zusammenhang mit dem von ihm vorgeschlagenen Konzept der "ökologischen Validität" sowie unter Bezugnahme auf ihn neuerdings Cicourel (13).

Was hier für die Methode des Experiments geltend gemacht wird, finden wir auch in theoretischen Überlegungen zu den Bestimmungsgründen sozialen Verhaltens, beginnend etwa in der Psychologie bei Lewin. Er plädiert dafür, das gesamte Handlungsfeld und Beziehungsfeld jedes konkreten Einzelfalls zu erfassen, um so über den Zwischenschritt der Bildung von Klassen ähnlicher Milieubedingungen oder Handlungssituationen zu verallgemeinerbaren Regelmäßigkeiten und zu quantitativen Aussagen vorstoßen zu können (14). Oder wörtlich: "Die Lebensnähe des Experiments ist nicht in der quantitativen Übereinstimmung mit der Wirklichkeit zu suchen; sondern entscheidend ist, ob beide Male wirklich derselbe Geschehenstypus vorliegt. Handelt es sich um

- (9) Gachowetz, Helmut: Feldforschung. In: Roth, E. (Hg.): Sozialwissenschaftliche Methoden, München, Wien, S. 263; ähnlich Bungard, Walter: Sozialpsychologische Forschung im Labor, Göttingen, Toronto, Zürich 1984, S. 20  
(10) Kreuz, Henrik, 1972: Soziologie der empirischen Sozialforschung, Stuttgart  
(11) Holzkamp 1964, a.a.O., S. 99  
(12) Brunswik, Eggon: Representative Design und Probabilistic Theory in a Functional Psychology. In: Psychological Review, 62/3, 1955, S. 193-217  
(13) Cicourel, Aaron V.: Interviews, Surveys and the Problem of Ecological Validity. In: The American Sociologist, 17, 1982, S. 11-20  
(14) Lewin, Kurt: Der Übergang von der aristotelischen zur galileischen Denkweise in Biologie und Psychologie. In: Grammann, C. F. (Hg.): Kurt-Lewin-Werkausgabe, Band 1, Bern, Stuttgart, 1981a, S. 233-278

Geschehnisse gleicher Struktur, so ist innerhalb breiter Bereiche ein Schluß... zulässig" (15).

#### Ähnliche Verallgemeinerungslogik im "Situationskonzept"

Fortwährende Situationsdefinitionen zwischen den Interagierenden stehen im theoretischen Zentrum des symbolischen Interaktionismus. Unter methodologischem Gesichtspunkt spricht Thomas (16) von der "Situationsmethode" als dem adäquaten sozialwissenschaftlichen Forschungswerkzeug. So wie der Experimentalforscher die Untersuchungssituation in seinem Design kontrolliert manipuliert und variiert, so soll der empirische Sozialforscher seine Daten gezielt in vielfältigen natürlichen und konkret beschreibbaren Situationen erheben (17). Das entspricht in etwa dem von BRUNSWIK vorgeschlagenen methodischen Vorgehen, wonach die in klassischen Designs übliche Isolierung von Variablen und das Sampling von Personen ersetzt werden sollen durch ein von ihm so genanntes "representative design", bei dem die in der Umwelt auftretenden Variablen in ihrer Zusammensetzung als Variablenpaket erhalten bleiben und zu diesem Zweck ein Sample von Situationen anstelle von Personen vorzunehmen ist (18). Ausdrücklich stehen solche Vorstellungen auch in BARKERS "ecological psychology" bei seinem "behavior setting"-Konstrukt im Vordergrund (19).

Unter Soziologen finden sich Anhänger des Situationskonzepts nicht nur bei Vertretern qualitativ orientierter Ansätze. K. R. POPPER etwa benutzt den Begriff "Situationslogik" für eine (allerdings nicht weiter ausgearbeitete) Methode, mit deren Hilfe menschliche Handlungen erklärbar und/oder verstehbar gemacht werden könnten (20). FRIEDRICHS (21) sowie FRIEDRICHS/LÜDTKE (22) schlagen explizit die Ziehung repräsentativer Situations-Samples vor. VASKOVICS (23) plädiert für sozialwissenschaftlich relevante Situations-Typologien. BUBA entwickelt Vorschläge dazu. Er sieht im Situationskonzept insbesondere Vorteile für die anwendungsorientierte Sozialforschung (um die es sich bei Begleitforschungen ja regelmäßig handelt): "Soll theoretische Überlegung und empirische Analyse sozialer Tatbestände in praktische Empfehlung und Beeinflussung der sozialen Realität münden, muß das theoretische Wissen so konkret sein, daß es in Handlungsanweisungen umsetzbar ist und eine Planungsgrundlage ... abgeben kann. Indem das Konzept Situation diesen hohen Grad an Konkretheit bietet, kann es Grundlage für eine Integration von Theorie und Praxis sein" (24). STEINERT (25) sieht gar

nach einer methodologischen Auseinandersetzung mit dem Interview als sozialer Interaktion das "Ende des Repräsentativitäts-Mythos" heraufziehen; er plädiert für eine Situations-Stichprobe, deren Repräsentativität nicht mehr generell, sondern nur noch themenspezifisch hergestellt werden könne.

Ähnliche Argumente vertreten andere Autoren im Zusammenhang mit dem Kontext-Begriff; am eindeutigsten wohl KRIZ, für den sozialwissenschaftliche Daten ohne Berücksichtigung ihres Entstehungs- und somit auch Geltungskontextes prinzipiell artefakt-verdächtig sind. "Ergebnisse empirischer Sozialforschung haben ... nur Sinn relativ zu einem Kontextrahmen, in dem die 'relevanten' Parameter des konstituierten Wirklichkeits-Segments angegeben und spezifiziert sind. Es sind jene Parameter, die für die Bildung phänomenologischer Äquivalenzklassen von 'vergleichbaren' Situationen pragmatisch bedeutsam sind, Situationen also, in denen aufgrund dieser Forschungsergebnisse Prognosen möglich sind und die Unsicherheit von Handlungskonsequenzen verringert wird" (26).

#### Beispiel Bildschirmtext

Wie sieht es nun mit der Übertragung solcher Überlegungen auf die Daten der Bildschirmtext-Begleitforschung aus? Anhand der Übersicht über Design-Strukturen bei Experimenten erkennen wir, daß wir es hier mit dem Fall "one-group pretest-posttest design" zu tun haben, kombiniert mit der Vorher-Messung aus dem Kontrollgruppendedesign. Allerdings existieren zwei einschneidende Abweichungen:

- Die Teilnahme am Feldversuch (Pilotprojekt) entschied sich (wie schon erwähnt) nicht nach dem Zufallsprinzip, sondern durch Selbstauswahl der Teilnehmer; bei der Kontrollgruppe aus der Wohnbevölkerung in Disseldorf/Neuss dagegen handelt es sich um eine nach dem Muster des Arbeitskreises Deutscher Marktforschungsinstitute (ADM) vorgenommene Haushalts-Repräsentativstichprobe. Im Unterschied zum Experimentaldesign wurde also nicht zunächst die gesamte Teilnehmerschaft an der Untersuchung rekrutiert und anschließend nach dem Zufallsprinzip auf Versuchs- und Kontrollgruppe aufgeteilt.

- Die Nullmessung in der Bevölkerung des Untersuchungsraums fand schon zwei Jahre vor der Nullmessung bei den Feldversuchsteilnehmern statt, was die Vergleichbarkeit der Daten einschränkt. Da sich jedoch am Informationsverhalten der Bevölkerung insgesamt in diesem Zeitraum nicht allzuviel geändert haben dürfte, ist diese zweite Einschränkung die weniger gravierende.

Alles in allem ist somit die Hauptuntersuchung der Bildschirmtext-Begleitforschung ein quasi-experimenteller Untersuchungsansatz, der am ehesten als (zudem unvollständiges) Design des Typs "nicht-äquivalente Kontrollgruppen" charakterisiert werden kann.

**Weiteres Fazit:** Für die Verallgemeinerbarkeit der Resultate über die Situation des Feldversuchs hinaus klingt dies alles (jedenfalls nach herrschendem Methodenverständnis) sehr schlecht!

#### Zum Vergleich: Kausalanalyse nicht-experimenteller Daten

Ein Blick in die Praxis gängiger und anerkannter empirischer Forschung (etwa zum Test von Hypothesen) zeigt jedoch, daß die Einschränkungen so ungewöhnlich gar

(15) Lewin, Kurt: Gesetz und Experiment in der Psychologie. In: *Grammann, aaO.*, 1981b, S. 315  
 (16) Thomas, William J.: Person und Sozialverhalten, Neuwied, Berlin 1965  
 (17) Vgl. auch die ausführliche Auseinandersetzung mit diesem Ansatz bei Volkart, Edmond H.: Soziale Verhalten und Situation. In: *Thomas 1965, aaO.*, insbes. S. 30ff.  
 (18) Brunswik 1955, aaO.; vgl. auch Kreppner, Kurt: Zur Problematik des Messens in den Sozialwissenschaften, Stuttgart 1975, S. 92  
 (19) Zum Beispiel Barker, Robert G.: Ecological Psychology: Concepts and Methods for Studying the Environment of Human Behavior, Stanford 1968  
 (20) Vgl. Beckmann, Jürgen: Logik der Situation. Begriffsanalyse und Probleme der empirischen Deutung, Bochum 1980 (Diplomarbeit, RUB)  
 (21) Friedrichs, Jürgen: Situation als soziologische Erhebungseinheit. In: Zeitschrift für Soziologie, 1974, S. 44-53  
 (22) Friedrichs, Jürgen; Lüdtke, H.: Teilnehmende Beobachtung, Weinheim 1971  
 (23) Vaskovics, Laszlo A.: Vorwort. In: Buba, Hans P.: Situation. Konzepte und Typologien zur sozialen Situation, Berlin 1980  
 (24) Buba 1980, aaO., S. 12  
 (25) Steinert, Heinz: Das Interview als soziale Interaktion. In: Meulemann, Heiner/K.-H. Reuband (Hg.): Soziale Realität im Interview, Frankfurt/M., New York 1984, S. 17-59

(26) Kriz, Jürgen: Die Wirklichkeit empirischer Sozialforschung. In: Bonif, W./H. Hartmann (Hg.): Entzauerte Wissenschaft, Soziale Welt, Sonderband 3, 1985, S. 77-89; ebenso ders.: Methodenkritik empirischer Sozialforschung, Stuttgart 1981

nicht sind; jedenfalls dann nicht, wenn man das Vorgehen hypothesentestender Forschung mit dem Idealdesign Experiment vergleicht (wie hier für die Begleitforschung geschehen).

Auf welche Art von Datenbasis greifen denn hypothesentestende Untersuchungen – also theoriegeleitete und nach den anerkannten Standards empirischer Sozialforschung verfahrenende Projekte – zurück? Man erhebt (in der Regel) zu *einem* bestimmten Zeitpunkt für eine repräsentative Stichprobe von Untersuchungseinheiten (Zufallsauswahl) einen möglichst umfassenden Katalog von Merkmalen, die sich auf die Dimensionen der zu prüfenden Hypothesen beziehen (abhängige, unabhängige, intervenierende Variablen). Um auf dieser Basis (nämlich nicht-experimentelle, deskriptive Survey-Daten) dennoch Hypothesen über Wirkungszusammenhänge (Kausalhypothesen) zu testen, muß man in der Phase der Datenanalyse das *rekonstruieren* versuchen, wozu man in der Phase der Datengewinnung nicht in der Lage war: Die nicht-experimentellen Daten werden so gruppiert, *als ob* sie im Rahmen eines Experimentaldesigns gewonnen worden wären; man versucht im nachhinein – wie Scheuch formuliert – „eine Umkehrung der Planung von Kontrollen beim Experiment“, womit jedoch allenfalls die Qualität eines Ex-post-facto-Experiments erreicht werde (27).

Typisches Beispiel dafür ist die Tabellenanalyse (doch gilt die gleiche Vorgehenslogik für multivariate Verfahren). Es werden – entsprechend den ermittelten Ausprägungen der unabhängigen Variablen (der „treatment“-Variablen) Kontrollgruppen gebildet. Im dichotomen Fall heißt das: Es wird eine Gruppe identifiziert, die dem angenommenen verursachenden Faktor ausgesetzt worden ist (analog zur Experimentalt Gruppe); die zweite Gruppe ist dementsprechend dem angenommenen Ursache-Faktor nicht ausgesetzt gewesen (analog zur Kontrollgruppe beim Experiment). So gruppiert, können die Daten der zur Quasi-Experimentalt Gruppe zählenden Personen hinsichtlich der abhängigen Variablen verglichen werden mit den Daten der zur Quasi-Kontrollgruppe zählenden Personen. Mit anderen Worten: Im Zuge der an Kausalhypothesen orientierten Auswertung nicht-experimenteller Daten wird im nachhinein das quasi-experimentelle Design „static group comparison“ rekonstruiert. Das aber ist nun in der Tat *noch weniger* als im oben geschilderten Fall der Bildschirmtext-Begleitforschung oder vergleichbarer anderer Begleituntersuchungen.

„Na, wenn schon“, könnte darauf erwidert werden. „Immerhin haben wir eine repräsentative Stichprobe, immerhin können wir unter Anwendung von Modellen der induktiven Statistik verallgemeinern!“ Leider ist auch dies – nimmt man es methodologisch genau – ein Fehlschluß. Thematisiert wird dabei nämlich nicht das Problem der Selbstauswahl der Untersuchungspersonen. Abgesehen von dem Fall, daß alle Personen der Grundgesamtheit *zwangweise* einem bestimmten Stimulus, einem bestimmten Maßnahmenprogramm ausgesetzt sind, hängt es natürlich von der Entscheidung jedes einzelnen ab, ob er sich einer bestimmten Situation aussetzt, ob er eine bestimmte Maßnahme in Anspruch nimmt (also z. B. sein Geld für ein bestimmtes Unterhaltungsangebot ausgibt oder eine bestimmte Schulform besucht usw.). Und damit stehen wir bei der nachträglichen Klassifizierung in Personengruppen, auf die sich die unabhängige Variable (treatment X) ausgewirkt hat, in der Tat wieder vor dem Problem der unabhängigen Variable nicht ausgewirkt hat, in der Tat wieder vor dem Problem der Selbstauswahl: Alle Personen, die in der Stichprobe als „betroffen von treatment X“

(27) Scheuch, Erwin K.: Entwicklungsrichtungen bei der Analyse sozialwissenschaftlicher Daten. In: König, R. (Hg.): Handbuch der empirischen Sozialforschung, Band I, Stuttgart 1973a, S. 161–237

identifiziert werden können (rekonstruierte Experimentalt Gruppe), haben dieses „treatment“ *freiwillig* gewählt.

Ein weiteres Problem kommt hinzu: Die übrigen Personen in der Stichprobe, die die Maßnahme nicht in Anspruch genommen haben, können dies aus unterschiedlichsten Gründen getan haben: von der bewußten Entscheidung gegen das Angebot über indifferentes Desinteresse bis zu dem Fall, daß man das Angebot gern angenommen hätte, wenn es bekannt gewesen wäre. Die (rekonstruierte) Kontrollgruppe ist also eigentlich gar keine. Oder methodologisch ausgedrückt: Wir haben es zum einen mit Selbstrekrutierten bei der Quasi-Experimentalt Gruppe und zum anderen mit einer nicht-äquivalenten (hinsichtlich der untersuchungsrelevanten Variablen sehr heterogenen) Kontrollgruppe zu tun – trotz Zufallsauswahl für die Gesamtstichprobe. Wir sind also noch viel schlechter dran als im hier geschilderten Begleitforschungs-Beispiel. Wer methodologisch rigoros argumentiert, müßte – wie Lazarsfeld – die Anwendung von Signifikanztests bei der Analyse von Kontingenztafeln ablehnen (28). Zur Ableitung kausaler Interpretationen aus nicht-experimentellen Daten müßte man korrekterweise zweistufig argumentieren:

a) Unter welchen Bedingungen kommt es dazu, daß sich Personen der Maßnahme X aussetzen? Und:

b) Falls sich Personen der Maßnahme X ausgesetzt haben: Wie wirkt sich dies dann (auf diese Personen) aus?

Das aber ist exakt das gleiche Fragepaar wie:

a) Unter welchen Bedingungen nehmen Personen oder Haushalte ein neues technisches System oder ein neues Produkt etc. in Anspruch? (Hier: Unter welchen Bedingungen nehmen Haushalte am Bildschirmtext-Feldversuch teil?) Und:

b) Wenn sie es in Anspruch nehmen: Wie wirkt es sich dann auf diese Personengruppen aus?

### Vergleich: Begleit- und Survey-Forschung

Daß die herkömmliche und akzeptierte Forschungspraxis nicht unproblematisch ist, macht allerdings die Situation für die Begleitforschung noch nicht besser – allenfalls tröstlicher. Die Frage bleibt, wie denn trotz nicht-repräsentativer Stichproben verallgemeinerbare oder zumindest teilweise verallgemeinerbare Aussagen zu finden wären.

Ich verweise dazu noch einmal auf das Experiment-Design. Ein wesentliches Charakteristikum insbesondere von Laborexperimenten ist es, daß in ihnen die Randbedingungen bestimmter Handlungs- oder Verhaltenssituationen kontrolliert werden (teilweise unter Inkaufnahme drastischer Vereinfachung dieser Situationen im Labor) und daß vergleichsweise wenig Wert auf die Repräsentativität der Experimentalt Teilnehmer gelegt wird. Für die Verallgemeinerbarkeit der Ergebnisse steht unter methodologischen Gesichtspunkt das Kriterium „externe Validität“ im Vordergrund, wobei insbesondere auf die Vergleichbarkeit der Experimentalsituation mit alltäglichen Lebens- und Handlungssituationen abgestellt wird.

In der *Methodologie* der empirischen Survey-Forschung dagegen steht nicht die Vergleichbarkeit von Datenerhebungs- und *Realisationen* im Vordergrund, sondern die formale Frage der Repräsentativität von *Untersuchungseinheiten*. Verantwortlich dafür ist die Prämisse: Wenn ich eine strukturtreue Abbildung aus der Grundgesamt-

(28) vgl. Schenck 1973a, aaO., S. 177

heit der *Untersuchungseinheiten* (z. B. Personen oder Haushalte) durch die Stichprobe realisierte, dann habe ich damit zugleich auch eine strukturtreue Abbildung aller Handlungs- und Lebenssituationen realisiert, in die diese Personen oder Haushalte eingebunden sind. Diese Prämisse läßt sich aber – wie wir gesehen haben – umkehren: Wenn die *Untersuchungssituationen* eine strukturtreue Abbildung der für die Fragestellung relevanten Situationen und Lebenszusammenhänge darstellen, dann haben wir damit zugleich auch eine repräsentative Stichprobe der insgesamt in solchen Situationen und Lebenszusammenhängen agierenden Personen, Haushalte, Institutionen etc. erfaßt.

Zur Illustration soll wieder das *Beispiel Bildschirmtext* dienen.

Im Falle von Bildschirmtext ist die relevante *Erprobungssituation* wie folgt charakterisierbar:

- Bildschirmtext konfrontiert die Versuchshaushalte mit einem für sie neuen Informationsinstrument sowie mit einer für sie ungewöhnlichen Logik bei der Informationsbeschaffung (vorherige Definition des Informationsbedarfs, aktive Selektion der benötigten Informationen, Suchbaumlogik).
- Bildschirmtext konfrontiert die Versuchshaushalte ferner mit einem interaktiven Telekommunikationsinstrument, mit dessen Hilfe eine Reihe von Erledigungen direkt von zu Hause (vom Heimterminal aus) abgewickelt werden können. Das System verknüpft die private Sphäre des Haushalts direkt mit hochtechnisierten und -rationalisierten außerprivaten Bereichen.
- Bildschirmtext muß sich als neu hinzukommendes technisches System neben den bisherigen Medien und Informationsquellen in den Alltag des Haushalts einfügen und dort seinen ihm zukommenden Platz finden (bisherige Informations-Quellen ergänzen, ersetzen, zusätzliche neue Funktionen übernehmen).

Übertragen auf Bildschirmtext, folgt aus den dargestellten Überlegungen die Behauptung:

Für eine Abschätzung der Auswirkungen der Nutzung des Systems sind die genannten Situationsseigenschaften von größerer Relevanz als die Merkmale der Haushalte und Personen, die das System nutzen (29).

Eine solche Behauptung muß natürlich *empirisch* Bestand haben, soll sie als Regel für die Übertragung von Resultaten der Nutzer auf die Nichtnutzer – d. h. als "Korrespondenzregel" für die Verallgemeinerung – brauchbar sein. Ihre Geltung würde bedeuten, daß die Nichtrepräsentativität der Stichprobe und daß das Auswahlprinzip "Selbstselektion der Versuchsteilnehmer" keine gravierenden verzerrenden Einflüsse ausübt, soweit es um die Beurteilung von Akzeptanz, von Nutzungsgewohnheiten und Nutzungsproblemen, von Auswirkungen auf andere Tätigkeitsbereiche u. ä. geht.

Mit anderen Worten: Daß die Struktur der im Pilotprojekt realisierten Stichprobe (freiwillige Teilnehmer am Feldversuch) ganz offensichtlich in einer Reihe von soziodemographischen Merkmalen von der Struktur der Wohnbevölkerung im Untersuchungsraum abweicht, ist kein Indiz für die Schwäche der Stichprobe. In den Abweichungen dokumentieren sich vielmehr Hinweise auf die Struktur der Grundgesamtheit, für die die Untersuchung sinnvolle Aussagen hervorbringen kann. Eine bevölkerungsrepräsentative Stichprobe dagegen ist dann *kein* Garant für die externe Validität

(29) Zur Klarstellung: Bildschirmtext wird hier lediglich als illustrierendes Beispiel benutzt, das durch andere Programm-Erprobungen ersetzbar ist.

empirischer Befunde, wenn nur ein systematischer Ausschnitt der Bevölkerung vom Untersuchungssachverhalt betroffen ist. So würde beispielsweise eine kontrollierte Zufallsauswahl mit Ausschaltung aller Selbstrekrutierungseffekte zwangsläufig *dann* Forschungserfakte produzieren, wenn in der Realität der Zutritt zum Nutzerkreis über den Marktmechanismus (Kauf eines entsprechenden Geräts) geregelt ist.

Daß Repräsentativstichproben im herkömmlichen Verständnis nicht in jedem Fall das Mittel der Wahl sind, ist keine neue Erkenntnis. So fragt etwa René König (30): "... was nutzt das präziseste Sample, solange nicht aus rein sachlogischen Erwägungen heraus gesichert ist, daß es das gesuchte Phänomen auch trifft?" Und E. K. Scheuch (31) bemängelt, unter Verweis auf entsprechende Warnungen von Zetterberg (32), daß die maßstäbliche Zusammensetzung der Auswahl zu gedankenlos als Kriterium guter Sozialforschung angesehen werde. Die statistische Repräsentativität einer Auswahl sei in manchen Fällen nicht nur ein *unzureichendes*, sondern sogar ein *falscher* Maßstab für die methodologische Beurteilung einer Forschung. Gerade ein im Vergleich zu Repräsentativitätskriterien "mißlungenes" Sample könne im Gegensatz zur verbreiteten Auffassung von erheblichem Erkenntniswert sein (33).

So zeigt etwa die aktuelle Entwicklung des Bildschirmtext-Dienstes der Bundespost: Die Schwierigkeiten, für das Pilotprojekt eine hinreichend große Zahl von Teilnehmern anzuwerben, finden ihr Spiegelbild in der sehr schleppenden Verbreitung des Systems seit der bundesweiten Einführung. Die skizzierte "Verzerrung" des Samples der Feldversuchs-Teilnehmer findet ihr Abbild in der Zusammensetzung des jetzt seit der Einführung des Systems sich bildenden Nutzer-Feldes.

Zurück zu der obigen These: Für eine Abschätzung der Auswirkungen der Nutzung des Systems sind die genannten Situationsseigenschaften von größerer Relevanz als die Merkmale der Haushalte und Personen, die das System nutzen.

Diese These hat sich nun in der Tat für den Bildschirmtext-Feldversuch bestätigt. In einem an der Ruhr-Universität zusätzlich zur Hauptuntersuchung der wissenschaftlichen Begleitung durchgeführten Projekt (34) wurden ergänzend zum Teilnehmerfeld der "Selbstmelder"-Haushalte aus den im Feldversuch unterrepräsentierten Bevölkerungsgruppen (untere Angestellte, Arbeiter) Haushalte gewonnen, die das Bix-Gerät leihweise kostenlos zur Verfügung gestellt erhielten. Schon nach einem relativ kurzen Erprobungszeitraum unterschieden sich deren Nutzungsgewohnheiten und deren Beurteilungen nicht mehr wesentlich von den Gruppen der unteren Angestellten und Arbeiter, die sich freiwillig und damit auf eigene Kosten am Feldversuch beteiligten (Selbstmelder).

Die spezifischen Situationsmerkmale hatten auch ein erheblich höheres Gewicht als die soziodemographischen Merkmale der Teilnehmerhaushalte und als die verwendeten Erhebungsinstrumente: Die empirischen Aussagen über Urteile und Nutzungsgewohn-

(30) König, René: Beobachtung und Experiment in der Sozialforschung, Köln, Berlin 1962, S. 25

(31) Scheuch, Erwin K.: Auswahlverfahren in der Sozialforschung, In: König, R. (Hg.): Handbuch der empirischen Sozialforschung, Band 3a, Stuttgart 1973b, S. 11

(32) Zetterberg, Hans: On Theory and Verification in Sociology, Totowa, N. J. 1954

(33) Scheuch 1973b, aaO., S. 12

(34) Jansen, Dorothea/H. Krommcy: Bochumer Untersuchung, Ergebnisbericht der Gruppendiskussionen, Band 5 der Reihe: Wissenschaftliche Begleituntersuchung Feldversuch Bildschirmtext Düsseldorf/Neues (hrsg. von der Landesregierung Nordrhein-Westfalen), Bochum, Düsseldorf 1983. Sowie: Krommcy, Helmut/D. Jansen u. a.: Bochumer Untersuchung, Systematische Inhaltsanalyse der Gruppendiskussionen, Band 11 der Reihe: Wissenschaftliche Begleituntersuchung, ... aaO., Bochum, Düsseldorf 1984



heiten auf der Basis der Daten der Bochumer Untersuchung (Gruppendiskussionen nur mit Haushalten der Kategorien untere Angestellte und Arbeiter) fanden ihre Entsprechung in den mit standardisierten Instrumenten erhobenen Daten der Begleitforschung für die Gesamtheit der Teilnehmer, in denen obere Angestellte und Selbständige besonders überrepräsentiert waren.

#### Empirische Identifikation "äquivalenter" Gruppen bei Btx-Teilnehmern und Bevölkerungstichprobe

Obwohl erfreulich, reicht dieses Zwischenfazit jedoch noch nicht für differenzierte Verallgemeinerungen. Denn bisher ist damit nur plausibel gemacht, daß für die spezifischen Fragen dieser Untersuchung das System Bildschirmtext und seine Nutzungsmöglichkeiten von größerem Gewicht sind als die üblicherweise herangezogenen sozialwissenschaftlichen Kontrollvariablen sozio-demographischer Art. Und: Dies gilt zunächst nur für die Teilnehmer an der Erprobung und liefert noch keinen Hinweis für Verallgemeinerungen über diesen Kreis hinaus auf die potentiellen Nutzer unter den bisherigen Nicht-Teilnehmern. Benötigt wird zusätzlich eine Antwort auf die Frage, wer diese potentiellen Nutzer sind und wie stark sie in der Bevölkerung vertreten sind.

Wenn sozio-demographische Merkmale für diesen Zweck kaum weitergehende differenzierte Schlußfolgerungen erlauben, dann wird dazu ein alternativer Satz von Variablen benötigt. Dieser Variablenatz muß für beide Populationen (Feldversuchsteilnehmer und Bevölkerung im Versuchsraum) verfügbar sein und zugleich eng genug mit dem Untersuchungsgegenstand zusammenhängen, um – wie KRIZ es ausdrückt – themenrelevante "Äquivalenzklassen" von vergleichbaren Situationen oder Kontexten rekonstruieren zu können.

Es gehört zu den sozialwissenschaftlichen Binsenweisheiten, daß Handlungen, Aktivitäten von Personen nicht isoliert und nicht einzeln für sich genommen existieren und Sinn ergeben, sondern daß sie eingebunden sind in allgemeinere Handlungsorientierungen, Verhaltensmuster, vielleicht sogar Lebensstile. Ziel statistischer Analysen (insbesondere multivariater Verfahren) ist es im Normalfall ihrer Anwendung, aus diesem Gesamtwirkungskomplex die Wirkungen einzelner Faktoren unter Kontrolle der anderen zu isolieren. Im gegebenen Fall stellt sich die Frage umgekehrt: Es gilt, einen Zusammenhangskomplex zu identifizieren und mit meßbaren Merkmalen zu beschreiben, in den auch die Nutzung des neuen Instruments (hier Bildschirmtext) eingebunden ist.

Ich will hier nicht im einzelnen auf kommunikationstheoretische Argumentationen eingehen, sondern mich auf den Hinweis beschränken: Unter Kommunikationsforschem gehört es ebenso zu den unbestrittenen Selbstverständlichkeiten, daß (zum einen) Kommunikation und Information wesentlicher Bestandteil jeder Art von Sozial- und Lebenszusammenhängen sind und daß (zum anderen) Informations- und Kommunikationshandlungen für die Analyse nicht aus ihren Zusammenhängen gelöst werden dürfen. Dementsprechend gilt: Die Nutzung eines neuen Kommunikations- und Informationsinstruments kann nicht als isolierte, nur auf dieses spezifische Medium beschränkte Aktivität verstanden werden. Sie ist – so darf unterstellt werden – eingebunden in ein generell aktiveres Informationsverhaltensmuster, in die Neigung, den eigenen Haushalt möglichst vollständig mit informationstechnischen Geräten, mit Informations- und Nachschlagewerken etc. auszustatten.

Akzeptiert man dies, so folgt daraus: Die Art und Stärke dieses Informationsstils (vielleicht sogar Informations-"Lebensstils") in der Bildschirmtext-Teilnehmerschaft

einerseits und in der Gesamtbevölkerung auf der anderen Seite wäre dann der "Transmissionsriemen" zur Übertragung der Pilotprojektbefunde auf die Nicht-Teilnehmer.

In der Tat bestätigten regressionsanalytische Auswertungen verschiedener Daten aus der Voruntersuchung (Bevölkerungstichprobe) mit Angaben zur Bereitschaft, ein neues Medium wie Bildschirmtext zu nutzen: Diejenigen, die die bisher schon verfügbaren Informationsmöglichkeiten am stärksten ausschöpften und die die vielfältigsten Informationsinteressen äußerten, waren auch am ehesten bereit, die Optionen eines zusätzlichen Informations- und Kommunikationsangebots zu nutzen.

Entsprechende Daten waren in großer Differenziertheit sowohl in der bevölkerungsrepräsentativen Untersuchung als auch in der Nullmessung der Btx-Teilnehmer erhoben worden. Sie liefern die empirische Basis für einen Vergleich beider Stichproben. Das entsprechende statistische Verfahren (35) sei hier nur kurz benannt: Getrennt voneinander wurden sowohl die Bevölkerungstichprobe als auch die Teilnehmer am Bildschirmtext-Feldversuch nach ihrem Informationsverhalten in homogene Gruppen eingeteilt. Dies geschah mit Hilfe einer *Clusteranalyse*, wobei als Indikatoren für Informationsverhalten die Angaben zur Häufigkeit der Informationsnachfrage in einer Vielzahl von Themenbereichen dienten.

#### Empirische Ergebnisse

Es zeigte sich, daß in der Tat in Bevölkerungstichprobe und Feldversuchsteilnehmerschaft identische *Informationsverhaltensmuster* existierten. Inhaltlich interpretiert erstreckte sich diese empirische Klassifikation von einer Gruppe mit "in jeder Hinsicht überdurchschnittlicher Informationsnachfrage" bis zu einer Gruppe mit "in jeder Hinsicht unterdurchschnittlicher Informationshäufigkeit".

In der folgenden Abbildung werden zur Illustration zwei von sieben ermittelten Informationsprofil-Gruppen aus der Btx-Teilnehmerschaft den entsprechenden Gruppen aus der Bevölkerung gegenübergestellt. Der besseren Übersichtlichkeit wegen sind die einzelnen Informationsarten, über die Daten (Informationsbedarf, Informationshäufigkeit) erhoben wurden, zusammengezogen worden: Kultur, Ratgeberinformationen zur Problem- und zur Alltagsbewältigung, Fahrplan- und Notdienstangaben usw. bis hin zum Sport. Es zeigt sich eine weitgehende Identität der unabhängig voneinander gewonnenen Klassifikationen in der Bevölkerungs- und in der Bildschirmtext-Stichprobe, mit nur geringen, jedoch inhaltlich interpretierbaren Abweichungen (36).

Die Existenz solcher nach ihrem Informationsverhalten äquivalenter Gruppen in der Bevölkerungstichprobe *und* in der Btx-Teilnehmerschaft ist durchaus keine Selbstverständlichkeit. Sie wäre dann nicht gegeben, wenn die Btx-Teilnehmer eine extrem selektive Auswahl von Personen mit sozusagen abweichendem Verhalten darstellen (nämlich abweichendem Informations-Verhalten: Technik-Freaks, grundsätzlich an allem Neuen Interessierte u. ä.). In diesem Fall müßten auch die Informationsnachfrage-Profile der Bevölkerungstichprobe von denen der Btx-Teilnehmer grundsätzlich verschieden sein. Eine Übertragung empirischer Befunde von der einen auf die andere Population wäre auf diesem Wege dann nicht möglich. Aber auch das wäre ein nicht unwichtiger Befund gewesen, der zumindest vor überreichten Schlußfolgerungen geschützt hätte.

(35) Dargestellt wird das Vorgehen in Kromrey, Helmut/H. Treinen: Die Verbreitung von Bildschirmtext in Privathaushalten. In: Rundfunk und Fernsehen 1985/2, S. 210ff.

(36) Für eine ausführlichere und alle Nachfrageprofile umfassende Darstellung s. Kromrey/Treinen 1985, aaO., S. 213ff.

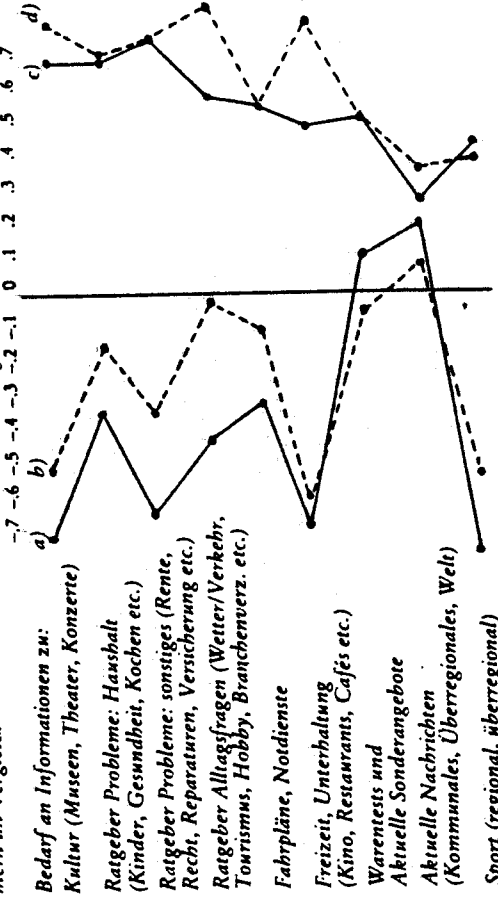
### Informationsnachfrage-Profile als Hinweis auf Ähnlichkeit von Lebenszusammenhängen

Darüber hinaus zeigt sich, daß die gefundenen Klassifikationen nicht nur als Informationsnachfrage-Profile, sondern als Ausdruck eines darüber hinausreichenden Lebenszusammenhangs zu interpretieren sind. Für die jeweiligen Gruppen in beiden Populationen gilt: Die Gruppenmitglieder sind hinsichtlich ihres allgemeinen Bezugs zur Technik, ihrer Tätigkeitsmerkmale, ihres Bezugs zur institutionellen sozialen Umwelt (z. B. Behörden und Organisationen) sowie auch hinsichtlich des Umfangs ihres Verkehrskreises untereinander ähnlich und zwischen den Gruppen deutlich unterschieden (37). Eine Übertragung der Befunde aus Btx-Gruppe 1 auf die Bevölkerungsgruppe 1, von Btx-Gruppe 2 auf die Bevölkerungsgruppe 2 usw. erweist sich damit für den hier untersuchten Zusammenhang als möglich und begründet. Da zudem auch die relativen Anteile der einzelnen Informationsverhaltens-Gruppen an der jeweiligen Gesamtheit der Fälle durch die Analyse bekannt sind, ist sogar eine quantitative Hochrechnung möglich.

### Klassifikation als Instrument der Generalisierung

Die damit angesprochene Verallgemeinerungslogik ist im übrigen nicht nur von den Anhängern des Experimentaldesigns und nicht nur von Verfechtern des Fallstudienan-

Abbildung: Informationsnachfrage-Profile von Bevölkerungsgruppen und Bildschirmtext-Teilnehmern im Vergleich



Nachfrageprofil: alle Informationsarten werden unterdurchschnittlich nachgefragt, außer Aktuelles

● a) Bevölkerungsguppe

● b) Btx-Teilnehmergruppe

● c) Bevölkerungsguppe

● d) Btx-Teilnehmergruppe

● Nachfrageprofil: alle Informationsarten werden überdurchschnittlich nachgefragt

Erläuterung: Der Nullpunkt des Index bezeichnet den Durchschnitt der Nachfragehäufigkeit nach einer Informationsart in der Gesamtheit der repräsentativen Bevölkerungsschichtprobe. In den Nachfrageprofile wird dargestellt, wie weit der Mittelwert der Nachfrage in einer homogenen Teilgruppe von Personen von diesem Gesamtdurchschnitt nach oben bzw. unten abweicht.

(37) Ausführlicher s. Kromrey/Trinen 1985, aaO., S. 218ff.

satzes in der qualitativen Forschung bekannt. Eine typische Aussage aus letzterer Richtung ist etwa: "Eine qualitative Methode, die ... rekonstruktiv verfährt, verallgemeinert aufgrund der Rekonstruktion von Strukturierungsgesetzmäßigkeiten, die bei einem Fall vorgefunden wurden. Nur wenn man bei anderen Fällen die gleiche Struktur expliziert hat, kann man davon ausgehen, daß diese durch die gleichen Strukturierungsmerkmale hervorgerufen wurde, und dann von einer Verallgemeinerung sprechen (38)." Diese Argumentation erinnert an die von Garfinkel entwickelte "dokumentarische Methode der Interpretation" mit ihren ständigen Wechselschritten zwischen Einzelfallanalyse und Herausarbeitung verallgemeinerbarer Muster (39).

Auch Vertreter der analytischen Wissenschaftslehre verweisen auf die Verwendung empirischer Typologien zum Zwecke der Verallgemeinerung. So schreibt beispielsweise Hempel: "Die Eigenschaften, welche zur Definition der verschiedenen Typen [verständen als Bündel von zusammen auftretenden Eigenschaften, H. K.] dienen, sollten nicht nur saubere Schubfächer zur Unterbringung der Einzelfälle des Untersuchungsbereichs liefern, sondern sie sollten auch einer vernünftigen Generalisierung dienen..." (40). Rosenmayr (41) spricht von "korrespondierender Typenbildung" als dem "tragenden Gerüst" bei der Interpretation empirischer Befunde; und bei Mangold (in seinem Lehrbuch zur empirischen Sozialforschung) finden wir: "Zur Beantwortung dieser Frage [Geltung über die Stichprobe hinaus, H. K.] muß man Überlegungen darüber anstellen, ob zwischen den untersuchten Fällen und dem interessierenden größeren Kollektiv strukturelle Übereinstimmungen bestehen, die als Basis für begründete Verallgemeinerungen dienen können (42)."

Obwohl also als Prinzip durchaus weit verbreitet, fehlt es an empirischen Bemühungen, die Leistungsfähigkeit einer strukturrekonstruierenden Verallgemeinerungsmethode abzuschätzen. Eine der sehr seltenen empirischen sozialwissenschaftlichen Studien zu dieser Frage stammt von Leslie (43). Er berichtet über eine Re-Analyse von - nach seinen Angaben - "Hundertern von Forschungsberichten", in denen auch auf die Auswirkungen des non-response, also der Nichtbeantwortung von Fragebögen, eingegangen wird, und zwar insbesondere bei postalischen Befragungen, die ja chronisch an zweifelhafter Repräsentativität wegen geringen und verzerrten Antwortrücklaufs leiden. Sein Ergebnis ist, daß selbst bei merklichem Bias - gemessen an den Verteilungen sozio-demographischer Standardvariablen - die themenbezogenen Ergebnisse zwischen Befragten, die geantwortet haben, und der Non-Response-Kontrollgruppe sich nicht

(38) Aufwanger, Stefan: Rekonstruktive Sozialforschung. In: "Jop", Universität Mainz, Ausgabe WS 1985/86, S. 7; ähnlich, Bude, Heinz: Die individuelle Allgemeinheit des Falls. In: Franz, H.-W. (Hg.): 22. Deutscher Soziologentag 1984, Beiträge der Sektions- und Ad-hoc-Gruppen, Opladen 1985, S. 86

(39) Vgl. Garfinkel, Harold: Common Sense Knowledge and Social Structure: The Documentary Method of Interpretation in Lay and Professional Fact Finding. In: J. M. Scher (ed.): Theories of the Mind, New York 1962, S. 689-712

(40) Hempel, Carl G.: Typologische Methoden in den Sozialwissenschaften. In: Tapscott, F. (Hg.): Logik der Sozialwissenschaften, Köln 1972, S. 86f. - Ähnlich mit seinem Konzept des "constructed type" Becker, Howard: Through Values to Social Integration, Essays on Social Contexts, Durham, N.C. 1950

(41) Rosenmayr, Leopold: Beitrag zur Podiumsdiskussion, in: Kausch, Max/W. Ott/E. K. Sibirsch (Hrsg.): Empirische Sozialforschung in der modernen Gesellschaft, Frankfurt/M., New York 1983, S. 87f.

(42) Mangold, Werner: Empirische Sozialforschung, Heidelberg 1972, S. 28f.

(43) Leslie, Larry L.: Are High Response Rates Essential to Valid Surveys? In: Social Science Research, 1972/1, S. 323-334

nennenswert unterschieden, *sofern* es sich bei den Befragten um homogene Teilpopulationen handelte – homogen hinsichtlich ihres Bezugs zur Untersuchungsfrage. Als Beispiel nennt er: Arbeiter beantworteten Fragen nach ihrer Beschäftigungssituation und ihrer Arbeitszufriedenheit oder ihrer Einstellung zu Gewerkschaft oder Unternehmensführung eher als Mitglied der Bezugsgruppe Arbeiter und nicht so sehr als z. B. männliche Person über 45 Jahre in der Einkommensgruppe 7000 bis 8000 Dollar. Solche (Bezugs-) Gruppenzugehörigkeiten würden andere persönliche und demographische Merkmale weit in den Schatten stellen.

### Resümee

Zusammenfassend läßt sich also feststellen, daß die von mir referierten Vorschläge auf eine Umkehrung empirischer Forschungslogik auf zwei Dimensionen hinauslaufen: nämlich hinsichtlich der Beziehung einerseits zwischen theoretischem Konstrukt und Indikator, andererseits zwischen Grundgesamtheit und Stichprobe.

*Zum ersten:* Das herkömmliche Verfahren der Operationalisierung eines theoretischen Konstrukts besteht in der Explikation seiner empirischen Geltung und der Benennung und Begründung von Indikatoren, die die theoretische Vorstellung meßbar machen sollen. Im Falle der Untersuchung in einem Feld, das sich noch in der Konstituierungsphase befindet (Erprobung neuer Organisationsformen oder neuer Technologien oder neuer Curricula), ist die Denkrichtung umgekehrt: Deskriptiv erfaßbar sind zunächst eine Reihe von Erscheinungsformen, deren Status als *Indikator für* ein übergreifendes Konstrukt erst noch zu rekonstruieren ist, oder: deren Geltungskontext erst noch erschlossen werden muß.

*Zum zweiten:* Das herkömmliche Verfahren der Stichprobenplanung besteht darin, zunächst eine empirisch abgrenzbare, zumindest im Erhebungszeitraum stabile Grundgesamtheit von Fällen zu definieren, für die die Aussagen der Untersuchung Geltung beanspruchen sollen (etwa wahlberechtigte Bevölkerung der Bundesrepublik im Juni 1987). In einem Stichprobenplan sind sodann die Regeln zu formulieren, nach denen eine repräsentative Teilmenge von Untersuchungsfällen gefunden werden kann. Im Falle der sozialwissenschaftlichen Begleitung von Feldversuchen und Modellprojekten konstituiert sich dagegen auf je spezifische Weise zunächst eine bestimmte Menge von Untersuchungseinheiten (also: Teilnehmer am Modellprojekt). Zu den Aufgaben des Forschers gehört es dann, die dazugehörige Grundgesamtheit zu rekonstruieren, für die die vorgefundene Menge von Fällen eine repräsentative Stichprobe ist.

Das ist im übrigen eine gar nicht so ausgefallene Situation, wie dies vielleicht auf den ersten Blick erscheinen mag. Kruskal und Mosteller – die ich ganz zu Beginn zitiert habe – weisen auch auf diesen Fall hin. Als Beispiel nennen sie die medizinische Forschung, wo es eher die Regel sei, daß der Forscher mit sich selbst bildenden Stichproben konfrontiert sei, für die er die zugehörige Grundgesamtheit zu bestimmen habe (44). Smith (45) verallgemeinert das Problem der "Samples in search of universes" auf alle Fälle abstrakter und schwer zugänglicher Grundgesamtheiten (etwa unverheiratete Mütter; oder ein aktuelles Beispiel bei uns: Volkszählungsgegner) sowie überhaupt auf ungenügend realisierte Stichprobenpläne.

Gerade letzteres ist in der jüngsten Vergangenheit sehr akut geworden. Wenn man sich die neuere Entwicklung der Antwortquoten bei Befragungsstudien ansieht, dann muß

(44) Kruskal/Mosteller 1979, III, 220f., S. 260f.

(45) Smith, Herman W.: *Strategies of Social Research*, Englewood Cliffs, N.J. 1973, S. 108ff.

sich der Eindruck aufdrängen, daß die hier am Beispiel der wissenschaftlichen Begleitung von Pilotprojekten aufgezeigte Schwierigkeit in Wirklichkeit ein generelles Problem der gesamten Survey-Forschung ist.

Wurden bei der Media-Analyse (einer allgemeinen Bevölkerungsumfrage, die mit sehr großem Aufwand versucht, eine möglichst hohe Ausschöpfung des Stichprobenplans zu erzielen) im Jahre 1980 noch 81 Prozent der Zielpersonen erreicht (46), so war die Ausschöpfungsquote bei dem im gleichen Jahr durchgeführten Allbus, für den aus finanziellen Gründen nur der forschungsübliche Aufwand betrieben werden kann (dreimaliges Nachfassen bei nicht erreichten Zielpersonen) lediglich 64 Prozent (47). Mittlerweile (Allbus 1986) ist sie sogar auf knapp 59 Prozent zurückgegangen (48).

Selbst wenn man unberücksichtigt läßt, daß schon der Stichprobenplan wegen der unzulässigen Informationsbasis über die Grundgesamtheit (seit 1970 keine aktuellen Volkszählungs-Daten) teilweise erhebliche Fehlgewichtungen enthält, kann bei so niedrigen Ausschöpfungsraten von einer repräsentativen Zufallsauswahl nicht mehr die Rede sein. Auch hier ist das Sample "auf der Suche" nach der zugehörigen Grundgesamtheit. Die kommerziellen Institute versuchen zwar, durch unterschiedliche Redressment-Verfahren die offensichtlichsten Mängel zu kompensieren, benutzen dabei allerdings sehr globale Merkmale wie Gemeindegrößenklassen, Bundesländergruppen, Haushaltsgrößen, Alters- und Geschlechter-Anteilswerte.

Es wäre also an der Zeit, systematisch und empirisch die Möglichkeiten der Generalisierung sozialwissenschaftlicher empirischer Befunde bei nicht-repräsentativen Stichproben zu erkunden, wo kontrollierte Zufallsauswahlen sich – aus welchen Gründen auch immer – nicht realisieren lassen.

(46) Verlagsguppe Bauer: MA '82. Reichweiten, Veränderungen, Trends; Hamburg 1982, S. 10

(47) Kirschner 1984, 220, S. 120

(48) Braun, Michael: *Erhebung, B. u.a.: Materialien zum Vortrag: "Der Allbus als Instrument der gesellschaftlichen Dauerbeobachtung"*, Mannheim 1987 (ZUMA, mimeo)