

Übersicht 38: Aufgaben des Pretests bei einer Inhaltsanalyse

- Umfang:** Einprozentige Stichprobe der geplanten Stichprobe von inhaltsanalytischen Erhebungseinheiten. (falls Material nicht bereits zugänglich): Darstellung der Untersuchungsziele.
 - Erhebungssituation:** Zeitpunkt für Analyse richtig gewählt? Material zugänglich oder noch in Herstellung (z. B. Zeitungen, Statistiken)?
 - Instrument:** Entspricht Untersuchungsplan den zu prüfenden Hypothesen? Sind die inhaltsanalytischen Einheiten eindeutig definiert? Kategorienschema anwendbar? Kategorienschema vollständig? Kategorienschema eindeutig in Zuordnungen? Differenzierung oder neue Kategorien erforderlich? Material vollständig? Erhebungseinheiten abgrenzbar? Periodisierung der Erhebungseinheiten angemessen? Mehrstufiges Stichprobenverfahren, Klumpenstichprobe oder einfache Zufallsstichprobe?
 - Kontrollen:** Vor allem durch Codierung gleichen Materials durch verschiedene Codierer. Reliabilität berechnen.
- 5.10.7. Fehlerquellen
- Die beiden hauptsächlichsten Fehlerquellen bei inhaltsanalytischen Studien sind die Reliabilität der Codierer (da man meist mit mehreren arbeitet) und die Validität des Kategoriensystems einschließlich der darauf aufbauenden Aussagen.
- Von den Codierern wird erwartet, daß sie ähnliche semantische Interpretationen der Wörter und Kategorien vornehmen. Die angestrebte Äquivalenz läßt sich einerseits durch Schulung, andererseits durch eine präzise Definition der Kategorien erreichen. Bei Kategorisierungen nach Intensität oder bei Bewertungsanalysen sollte die Einstellung der Codierer zum Objekt untersucht werden, um nicht zusätzliche Verzerrungen zu erhalten. Da beim Codierer Lerneffekte auftreten, kann man die ersten 10 Prozent des Materials nochmals codieren lassen, um diesen Effekt zu kontrollieren. Nach den vorliegenden Erfahrungen ist die Reliabilität des Codierers um so geringer:
- je höher die Zahl der Kategorien ist,
 - je weniger erschoßend die Kategorien definiert sind,
 - je mehr Bewertung des Kontextes erforderlich ist.
- Sie variiert (SPIEGELMAN, TERWILLIGER & FEARING 1953):
- mit der Mehrdeutigkeit des Materials,
 - mit der Mehrdeutigkeit der Codieranweisungen,
 - mit Fähigkeit und Training der Codierer.

Zum Training der Codierer geben BUDD, THORP & DONOHUE (1967, S. 52 ff.) ausführliche Hinweise.

Die Reliabilitätsmessungen haben die gleichen Formen wie die im Abschnitt 5.7.8 vorgeschlagenen. Man bedient sich der Einfachheit halber beim Vergleich zweier Codierer als Reliabilitätsmaß folgendes Quotienten: Anzahl der von beiden gleich codierten Wörter dividiert durch die Summe der von beiden insgesamt codierten Wörter. Kompliziertere Reliabilitätsmaße sind von KAPLAN & GOLDBEN (1949) und SCOTT (1955) entwickelt worden.

Die Validität der Studien beruht häufig nur auf ihrer Plausibilität. Es wurde bereits im Abschnitt 5.10.3 ausgeführt, daß die face-validity vor allem bei Studien über den latenten Inhalt von Kommunikationen unzureichend ist. Es bedarf entweder der Außenkriterien oder der Ableitung von Prognosen, um die Aussagen zu validieren. Eine Konstrukt-Validität wird nur in seltenen Fällen erreichbar sein (vgl. das Beispiel bei HOLSTRI 1968, S. 662).

Für die Validierung durch Außenkriterien (concurrent validity) ist die Studie von HOLSTRI & NORRIN ein gutes Beispiel. Untersucht wurden die politischen Ereignisse im Sommer 1914 vor Beginn des Ersten Weltkrieges. Die Autoren gingen von der Annahme aus, nicht nur die politischen, sondern auch die finanziellen Eliten würden auf die sich zuspitzende Krise reagieren. Es wurden a) die Einstellung der politischen Eliten aufgrund von Dokumenten analysiert und b) finanzielle Transaktionen parallel zur politischen Entwicklung erhoben. Die Korrelation der Maße aus der Inhaltsanalyse mit den finanziellen Indizes betrug .90, womit die Validität der Aussagen der Inhaltsanalyse über den Verlauf der Krise an Hand von Außenkriterien geprüft war (vgl. HOLSTRI 1969, S. 146 f.).

5.11. EXPERIMENT

5.11.1. Voraussetzungen

In der Umgangssprache versteht man unter Experiment Handlungen, durch die bestehende Bedingungen verändert werden, mit denen man etwas «ausprobiert» und nun abwartet, wie sich diese Veränderungen auswirken. In wissenschaftlichem Sprachgebrauch ist dieses Verständnis zu weit, da z. B. die Ungewißheit des Ausgangs Kennzeichen jeder Methode ist. Das wissenschaftliche Experiment kann definiert werden als «wiederholbare Beobachtungen unter kontrollierten Bedingungen, wobei eine (oder mehrere) unabhängige Variable(n) derartig manipuliert wird (werden), daß eine Überprüfungsmöglichkeit der zugrunde liegenden Hypothese (Behauptung eines Kausalzusammenhanges) in unterschiedlichen Situationen gegeben ist» (ZIMMERMANN 1972, S. 37).

Die Definition von ZIMMERMANN ist ein Versuch, die unterschiedlichen methodologischen Anforderungen zu berücksichtigen, denen ein Experiment

ment unterliegt. (Die Anforderungen sind keineswegs in allen experimentellen Untersuchungen erfüllt, obgleich auch sie als Experimente bezeichnet werden.) Die **Definition** ist daher strenger im Vergleich zu anderen, etwa der gängigen, ein **Experiment sei eine «Beobachtung unter kontrollierten Bedingungen»**. Sie hat zudem den Vorzug, sich nicht nur auf das Experiment als eine **spezielle Methode** der Sozialforschung anwenden zu lassen, sondern auch für das Experiment als eine spezifische **Form des Untersuchungsplanes** – ungeachtet der Methode – zu gelten.

Das Experiment gilt als die exaktere Form wissenschaftlicher Forschung, weil es folgende Vorteile bietet:

1. Kontrolle aller zur Hypothesenprüfung bedeutsamen Variablen.
2. Manipulation der Versuchsbedingungen, um den Einfluß einer oder mehrerer unabhängiger Variablen auf eine abhängige Variable festzustellen; Kausalanalyse.
3. Messung der Art des Zusammenhangs zwischen zwei Variablen, möglichst durch Angabe der mathematischen Funktion, die den Zusammenhang beschreibt.

Zu den Vorteilen des Experiments gehört auch, daß es nicht denkbar ist ohne eine **enge Verbindung von Hypothesen, Untersuchungsplan und mathematisch-statistischer Auswertung** – jener Verbindung, die für alle Methoden der Sozialforschung gefordert wird, hier aber in ihrer stringentesten Form auftritt.

Die Voraussetzungen, um die angeführten Vorteile exakter Hypothesenprüfung zu erreichen, liegen in der komplizierten methodologischen Struktur experimenteller Methodik. Die Kenntnis der wichtigsten Voraussetzungen ist zugleich ein Ausgangspunkt für die Frage, ob Experimente in der Soziologie möglich sind.

Ständig relevanten Themen

1. Wiederholbarkeit: Die experimentellen Bedingungen sollen nicht einmalig sein; Um zu prüfen, ob ein Film mit aggressiven Stimuli Personen aggressiver macht, werden zwei strukturell gleiche Personengruppen vor der Präsentation des Films auf den Grad ihrer momentanen Aggressivität hin untersucht. Der einen Gruppe wird dann ein Film über einen Boxkampf vorgeführt, der anderen Gruppe nicht. Nach der Vorführung des Films werden in beiden Gruppen nochmals Messungen des Grades der Aggressivität vorgenommen. Diese Erhebungssituation ist wiederholbar; fraglich ist, ob die Personen sich durch das Experiment so verändert haben, daß sie erst nach Wochen oder gar nicht mehr in einem solchen Experiment die gleichen Reaktionen zeigen, d. h. das Experiment zu den gleichen Ergebnissen führt. (Von denselben Ergebnissen die meisten makrosoziologischen Prozesse, z. B. eine veränderte politische Situation durch internationale Verträge oder soziale Veränderungen durch neue Gesetze.)

2. Kausalanalyse: Das zentrale Problem des Experiments ist, die **Beziehung zwischen den wenigen einbezogenen Variablen zu präzisieren**, da es

festlegung und Stärke der kausalen Beziehung der zu messen oder Variablen!

darum geht, die **Variation einer abhängigen Variablen durch eine oder mehrere unabhängige Variablen zu erklären**. Im Experiment wird, um eine Kausalbeziehung zwischen unabhängigen und abhängigen Variablen zu prüfen, eine asymmetrische Beziehung vorausgesetzt. D. h., nach dem in Abschnitt 2.5 aufgeführten Kriterien der Beziehungen zwischen Variablen: die Beziehung muß irreversibel sein; die Frage, ob es sich um eine hinreichende oder bedingte und notwendige oder substituierbare Beziehung handelt, bleibt dann im Experiment zu prüfen.

Um die Art der Kausalbeziehung zu prüfen, hat John Stuart Mill in seinem «System of Logic» fünf mögliche Formen kausaler Analyse entwickelt, von denen hier zwei besonders wichtig sind:

1. Methode der Übereinstimmung: ein Ereignis (abhängige Variable) tritt unter zwei Bedingungen auf. Unterscheiden sich diese Bedingungen bis auf eine Variable in allen anderen, d. h. sind alle restlichen Variablen unter der ersten Bedingung anwesend, unter der zweiten hingegen nicht, so ist die eine in beiden Bedingungen anwesende Variable die «Ursache» des Auftretens der abhängigen Variablen.

2. Methode der Differenz: In zwei Bedingungen tritt unter der ersten Bedingung ein Ereignis (abhängige Variable) auf, unter der zweiten Bedingung nicht. **Finden die Bedingungen in ihren Variablen bis auf das Fehlen einer Variable im zweiten Fall identisch, so gilt diese als «Ursache» für das Auftreten der abhängigen Variablen in der ersten Bedingung.**

Die Logik von Mill auf die Analyse sozialwissenschaftlicher Experimente anzuwenden unterliegt drei erheblichen Einschränkungen:

1. Mill setzt eine deterministische Beziehung zwischen den Variablen voraus, die meisten sozialwissenschaftlichen Gesetzesausagen sind aber statistische Gesetze (vgl. Abschnitt 2.2.1). Oft ist die unabhängige Variable X eine notwendige, aber nicht hinreichende Bedingung für das Auftreten der abhängigen Variablen Y.
2. Mill geht von didotomen Variablen aus (tritt auf – tritt nicht auf), während die meisten Variablen in den Sozialwissenschaften mehrere Ausprägungen haben, z. B. Alter, Beruf, Grad der Aggressivität.
3. Die Kausalbeziehung muß nicht auf eine der einbezogenen Variablen, sondern kann auf eine nicht gemessene weitere Variable zurückgehen. Das führt zu der nächsten Voraussetzung, der Kontrolle der Bedingungen.

3. Kontrolle der Bedingungen: Die Kontrolle der zur Hypothesenprüfung relevanten Bedingungen des Experiments setzt voraus, explizite oder implizite Gesetzesausagen über den Zusammenhang mehrerer Variablen anzuhängen. Solche Gesetzesausagen sind erforderlich, um zu begründen: a) welche unabhängigen Variablen eine Variation der abhängigen Variablen hervorrufen können (Auswahl der Variablen); b) welche interaktiven Effekte und Kovarianzen unter den unabhängigen Variablen auftreten können. Muß der Forscher in dem unter (1) genannten Aggressionsexperiment auch die folgenden Variablen kontrollieren: Alter der Personen, Va-

riablen einer schichtspezifischen Form der Reaktion auf verbale und physische Aggression oder das Ausmaß, in dem bei den am Experiment beteiligten Personen durch aggressive Modelle im Fernsehen die kulturelle Norm (hier eine Brutaliätsschwelle) herabgesetzt ist?

Die Kontrolle der Bedingungen im Experiment ist demnach eine Funktion des Standes der Theorie einer Disziplin. Ein weiteres, gerade für soziologische Experimente wichtiges Argument nennt NAGEL (1961, S. 451): «Die Ausübung von Macht, um soziale Bedingungen entsprechend den experimentellen Zwecken zu modifizieren, ist eindeutig selbst eine soziale Variable.» Daß es dem Forscher gelingt, Personen in sein Labor zu holen und zur Mitarbeit zu bewegen, zehrt von dem Vertrauen in die Ethik, den Sinn und den Nutzen der Forschung – ebenso wie bei anderen Methoden. Darüber hinaus entsteht das **Problem, daß sich weder eine Wirtschaftskrise** **Straf** **noch die Reaktion von Menschen auf ein Feuer im Büro durch ein Experiment erzeugen lassen**, sowohl aus Gründen der Macht wie denen der Ethik. Die Euthanasie-Experimente im nationalsozialistischen Deutschland zeigen jene Korruption wissenschaftlicher Ethik durch die Delegation von Macht an Wissenschaftler für ihre Experimente.

4. Situationen: Um kausale Beziehungen unter kontrollierten Bedingungen festzustellen, bedarf es einer **definierten Experimentalsituation** (Erhebungssituation). Der Forscher kann sie entweder herstellen oder aufsuchen, was vereinfacht der Unterscheidung von **Laborexperiment** und **Feldexperiment** entspricht.

5. Ethik: Ein weiteres Problem, das man den Voraussetzungen zuordnen muß, ist die Ethik der experimentellen Methode. In einigen Experimenten werden die Teilnehmer Ängsten ausgesetzt, sie sollen Elektroschocks verteilen oder werden mit eintönigen Aufgaben (z. B. lange Additionen) beschäftigt. In vielen Experimenten werden die Teilnehmer bewußt getäuscht, um die angezielten Reaktionen zu messen. Gerade **um der angestrebten Authentizität der Experimentalsituation willen sind solche Manipulationen erforderlich. Das Problem der ethischen Legitimation** verschärft sich noch durch die Anwendung experimenteller Methoden auf Gebiete wie Sexualität oder Drogenkonsum und allgemein bei allen Feldexperimenten (vgl. das Beispiel im Abschnitt 5.11.4).

Als minimale Anforderung an Experimente gilt, den Teilnehmern keinen Schaden zuzufügen, sie nachträglich über das Experiment aufzuklären und zu prüfen, ob das Experiment weitere Folgen für sie hatte. Einen generellen Code ethischer Regeln hat die American Psychological Association nach vierjähriger Forschung im Jahre 1971 vorgelegt. In ihm wird u. a. die Bezeichnung «Versuchsperson» durch «Teilnehmer» ersetzt, ein höherer Grad von Anonymität bei Beobachtungen gefordert, zu Studiengeräten, die für die Teilnehmer eine «kompensatorische, nutzbringende Erfahrung» sein sollen, und schließlich die Verpflichtung der Psychologiestudenten, an Experimenten teilzunehmen, (etwas) eingeschränkt. In noch

umfassender Weise dokumentiert die enorme Materialsammlung von Katz (1972) an Fällen aus allen Disziplinen die ethischen und juristischen Probleme des Experiments.

Die Probleme der Ethik sind letztlich nicht nur von der Chance der Teilnehmer, ein Experiment abzulehnen, oder von dem Verantwortungsbewußtsein des Versuchsleiters abhängig, sondern auch davon, ob das Vertrauen der Teilnehmer in eine Profession durch deren Ziele gerechtfertigt ist. Das emanzipatorische Potential der Wissenschaft, die Frage nach den Interessen der Betroffenen und die mögliche Praxis bilden die Bezugspunkte auch für experimentelle Forschung (vgl. hierzu Horzka 1970).

Aus der Schwerwichtigkeit in sozialwissenschaftlichen Untersuchungen alle genannten Voraussetzungen zu erfüllen, hat NAGEL (1961, S. 450 f.) den Schluß gezogen, neben dem kontrollierten Experiment eine weniger strenge Form zuzulassen, die «kontrollierte Untersuchung». Sie hat eine ähnliche methodologische Struktur wie das Experiment, richtet sich ebenfalls darauf, empirisch Gesetze zu prüfen, verlangt aber weder eine Manipulation der Variablen durch den Forscher noch die Reproduzierbarkeit der Erhebungssituation. NAGEL verweist in diesem Zusammenhang auf die Astronomie und die Astrophysik: Auch dort ist es nicht möglich, Variablen zu manipulieren und Erhebungssituationen herzustellen; das schließt aber eine kontrollierte Untersuchung der Planetenbahnen, die zu Gesetzesaussagen führt, nicht aus.

Experimentelle Verfahren sind ein zentraler Bestandteil der psychologischen und sozialpsychologischen Forschung, sehr selten hingegen einer der Soziologie. Hierfür gibt es zwei Erklärungen: **die Voraussetzungen des Experiments sind in der Soziologie (und Ökonomie) nicht gegeben, und:** der Entwicklungsstand der Disziplinen erlaubt noch keine so rigide Prüfung ihrer Theorien. Die erste Erklärung stellt sich bei näherer Prüfung methodologisch als unzutreffend heraus (Opp 1970 a).

Gegen Experimente in der Soziologie wird **zumeist die Künstlichkeit der Experimentalsituation und die Komplexität der sozialen Realität angeführt: Künstlichkeit** ist indessen kein notwendiges Kennzeichen von Experimenten; vielmehr läßt sich durch eine systematische Variation der Experimentalsituation eine höhere Gültigkeit der Experimente erreichen (Opp 1970 a, S. 41 f.). Gegen das Argument der Künstlichkeit spricht ferner: Die Situation im Labor unterscheidet sich von der Situation im Feld nicht prinzipiell, sondern nur graduell, nämlich durch die Zahl der Restriktionen und die Art, wie gut die strukturellen Elemente des Feldes abgebildet sind. Einige Experimente lassen sich gar nicht mehr eindeutig der Trennung Labor – Feld zuordnen, wie z. B. das im Abschnitt 5.11.4 genannte von Clark III & Word (1970). Problematischer ist die Angemessenheit der Abbildung: Zu Recht bezweifeln MAYNITZ, HOLM & HÜBNER (1971, S. 185), ob «das Lob des Experimentators oder der Gewinn von Pluspunkten, Spiel-

marken oder geringfügiger Geldbeträge für die Versuchspersonen [nicht] eine ganz andere (vermutlich geringere) Bedeutung besitzen als das Lob eines wirklichen Vorgesetzten, eine gute Examennote oder eine Gehaltssteigerung. Dies zu unterscheiden kann selbst Aufgabe von Experimenten oder Forschungen mit anderen Methoden sein.

Das zweite Hauptargument, die *Komplexität der sozialen Realität* (z. B. Siebel 1965, S. 185 ff.), ist kein spezieller Einwand gegen die Methode des Experiments, sondern auch gegen alle anderen Methoden der Sozialforschung, die doch gemeinhin von den Kritikern akzeptiert werden. Zudem läßt sich die Realität mit einer finiten Menge von Variablen (und Gesetzesaussagen) hinreichend beschreiben (Opp, 1970 a, S. 42 f.). Dieser Gegenstand von Opp ist methodologisch korrekt, nur praktisch häufig unzutreffend. Gerade um die Komplexität abzubilden, bedarf es entweder sehr komplizierter, multifaktorieller Untersuchungspläne oder eines außerordentlich langen Prozesses kumulativer Forschung aus Experimenten mit jeweils wenigen Variablen. Da der zweite Fall in der psychologischen und sozialpsychologischen Literatur überwiegt, entsteht dort die Schwierigkeit, die Einzelbefunde in eine Theorie größerer Allgemeinheit zu integrieren. Wie die lerntheoretische Forschung zeigt, ist das sehr wohl möglich. Hingegen haben die Forschungen zur Aggression und den Prozessen kognitiven Gleichgewichts resp. der kognitiven Dissonanz einen solchen Grad an Integration nicht erreicht – obgleich Soziologen gern die Theorie der kognitiven Dissonanz als Beispiel für eine entwickelte Theorie anführen. Zutreffender ist es, sowohl im Falle der Aggressions- wie der Dissonanzforschung von einer zunehmenden Differenzierung und Präzisierung der ursprünglichen theoretischen Begriffe zu sprechen. Dies hat zu einer Vielzahl von Experimenten geführt, die ihrerseits eine weitere Differenzierung der Variablen und Versuchsbedingungen nach sich zogen. Damit hat sich der Realitätsgehalt der Hypothesen erhöht; allerdings erschwerte die Interdependenz von Problemen der Konzeptualisierung, der Meßverfahren und der Anforderungen experimenteller Exaktheit – trotz der hohen methodischen Standards – die theoretische Integration der Einzelergebnisse.

Das faktische Dilemma benannte Blalock (1973, S. 2) treffend:

«But, unfortunately, we sometimes believe (or at least act) as though extremely complex 'real world' social systems can be studied as simply as laboratory situations. Of course we will know that this is not the case, but we collect and analyze our data as though it were! ... But if reality is complex, so must be our analysis. If there is any one thing that I have learned over the past dozen years or so, it is this fact. If, for example, a complex situation should require a model involving 30 variables reciprocally interrelated in a complex causal model, then it is wrong to attempt to break them apart into sets of two or three variables unless very restrictive assumptions hold true.»

Will man im Rahmen einer um strenge Hypothesenprüfung bemühten

soziologischen Forschung stärker als bisher mit experimentellen Verfahren arbeiten, so entstehen noch größere Probleme als z. B. in der Psychologie und Sozialpsychologie, die **komplexen Variablenzusammenhänge nicht in zu engen Experimenten mit nur wenigen Variablen zu prüfen. Erschwert wird durch, daß sich außerdem die oftmals unpräzise Formulierung der Hypothesen aus (die zweite obengenannte Erklärung).**

Hovland (1972) untersuchte einige Unterschiede in den Ergebnissen von Experimenten und Befragungen zur Massenkommunikationsforschung und zum Einstellungswandel. Er sieht in den Unterschieden keine Widersprüche, sondern führt sie auf verschiedene Bedingungen (Art der untersuchten Kommunikation, Art des kommunizierten Inhalts) zurück. Um einen höheren Grad an Integration und Validierung der Ergebnisse zu erreichen, plädiert er für komplexere experimentelle Untersuchungspläne, die sich an Hypothesen orientieren, die in Panel-Studien gewonnen wurden.

Die Verwendung von umfangreichen Kausalmodellen, von quasi-experimentellen Untersuchungsplänen, von fantasievollen Feldexperimenten und punktuellen Laborexperimenten dürfte die Möglichkeit bieten, jene Komplexität zu erhalten und gleichzeitig eine methodologisch und methodisch exakte Prüfung der Hypothesen zu erreichen.

5.11.2. Varianten

Das **Laborexperiment** wird von Festinger (1953, S. 137) definiert als ein Experiment, **«in dem der Forscher eine Situation mit genau den Bedingungen schafft, die er haben möchte und in der er einige Variablen kontrolliert und andere verändert»**. Das Experiment soll eine möglichst vollständige Kontrolle aller unabhängigen Variablen unter Ausschluß aller störenden Einflüsse enthalten. Auch die Zuweisung der Teilnehmer zu den unterschiedlichen Versuchsbedingungen ist nur im Laborexperiment exakt zu erreichen. Daß ein «reines» Experiment allein aufgrund der Einflüsse des Versuchsteilers nicht möglich ist, dürfte inzwischen erwiesen sein (vgl. Bredekamp 1969, S. 335 ff.).

Es geht darum, auch seine Einflüsse zu kontrollieren. Umstritten ist, ob das Laborexperiment Situationen außerhalb des Labors abbilden soll, was z. B. Festinger (1953, S. 137) zugunsten der Reinheit des Experimentes ablehnt.

Demgegenüber ist das **Feldexperiment** nur durch die Authentizität der Erhebungssituation gekennzeichnet. Es ist **«eine Untersuchung in einer realistischsten Situation, in der eine oder mehrere Variablen vom Versuchsleiter manipuliert werden, so sorgfältig, wie es die Situation gestattet»** (Kerlinger 1964, S. 382). Der zweite Teil des Satzes deutet bereits auf die Schwierigkeiten hin: Im Feldexperiment lassen sich faktisch nur ein

Teil der unabhängigen Variablen kontrollieren. Die Anfälligkeit des Feld-experiments gegen Störeffekte, nichttenbezogene Variablen und nicht-antizipierte Konsequenzen ist größer als beim Laborexperiment. Der geringeren Exaktheit steht eine größere Realitätsnähe gegenüber: Es bietet den Vorteil, komplexe Interaktionen untersuchen zu können, außerdem stärkere Effekte der unabhängigen Variablen auf die abhängige zu erreichen.

Quasi-experimentelle Verfahren: Unter dieser Bezeichnung fassen CAMREBELL & STANLEY (1963, S. 206) alle experimentellen Untersuchungspläne zusammen, in denen nur Teile der Anforderungen an Labor- und Feld-experimente erfüllt sind. Der Forscher hat nur eine unvollständige Kontrolle über die experimentellen Bedingungen: Es fehlt eine zufällige Zuweisung der Teilnehmer auf die einzelnen Bedingungen, oder es besteht Unsicherheit über den Zeitpunkt der Ersmessung vor Einführung eines experimentellen Stimulus oder Unkenntnis über die Person, die die erste Messung vornahm. Die Autoren diskutieren zehn solcher quasi-experimentellen Untersuchungspläne, durch die auch unter eingeschränkten Bedingungen eine möglichst exakte Hypothesenprüfung zu erreichen ist. Diese Verfahren kommen dem Vorschlag NAGERS, der «kontrollierten Untersuchung», nahe; gerade für die soziologische Forschung sind sie bedeutsam. Im Rahmen dieser Einführung können nur zwei dieser Pläne (Abschnitt 5.11.4) kurz behandelt werden; ansonsten sei auf den genannten Text und die Darstellung bei ZIMMERMANN (1972) verwiesen.

Ex-Post-Factor-Experiments. In dieser Form des «Experiments» wird eine Analyse vorliegenden Materials, z. B. von Daten aus einer Befragung oder Akten, vorgenommen, um den Einfluß eines früheren Ereignisses auf die Entwicklung der Personen zurückzuverfolgen.

So hat CHRISTIANSEN (zit. nach CHAPIN 1956, S. 238 ff.) die Hypothese «Je größer der Fortschritt eines Schülers an der Höheren Schule ist, desto größer wird der Grad seiner wirtschaftlichen Anpassung sein» an einer Stichprobe von rund 2000 Schülern aus dem Jahre 1926 ex post im Jahre 1935 untersucht. Aus den Akten war erkennbar, welche Schüler die Schule mit und welche sie ohne Abschlußzeugnis verlassen hatten. Die Schüler wurden, soweit noch erreichbar, aufgesucht und interviewt. Aus der Analyse der Daten der Interviews konnten aus beiden Gruppen von Schülern jeweils 145 ausgewählt werden, die sich in sechs Variablen gleich waren: Durchschnittsnoten, Beruf des Vaters, Alter, Geschlecht, sozialer Status der Nachbarschaft, Nationalität der Eltern. Der Vergleich der beiden Gruppen erbrachte, daß von der Kontrollgruppe vorzeitig Abgegangen 83% seit dem Abgang ein gleich hohes oder höheres Gehalt bezogen (Indikator der wirtschaftlichen Anpassung), während es in der Experimentalgruppe der Absolventen 89% waren – ein sehr geringerer Unterschied also.

Hier wurde eine differenzierte Kontrolle der Variablen vorgenommen, um die Gruppen bis auf die abhängige Variable möglichst zu parallelisieren. Handelt es sich dabei um ein Experiment? – Entgegen GREENWOOD (1956, S. 182) muß dies verneint werden, da im strengen Sinne keine Kontrolle

aller unabhängigen Variablen besteht und von nur einer Messung der Effekte ex post auf die Ursache (unabhängige Variable) geschlossen wird. ZIMMERMANN (1972, S. 186) spricht daher von einer «Anordnung». Dennoch ist diese Anordnung eine brauchbare und heuristisch wertvolle Form der Analyse von Variablen-Zusammenhängen.

Simulation, Reale Prozesse zu simulieren heißt, sie in einem Modell nachzuahmen. Aufgrund vorhandener Daten werden soziale Prozesse reproduziert, was unter der Voraussetzung, über eine möglichst exakte Kenntnis der Beziehungen zwischen den Variablen zu verfügen, eine systematische Variation der Effekte der unabhängigen Variablen auf die abhängige(n) Variable(n) ermöglicht. Im Bereich makrosoziologischer und makroökonomischer Forschung kann dies z. B. zur Stadt- oder Regionalplanung geschehen. Die Möglichkeiten zur Simulation sind vor allem durch die Verwendung von Computern gestiegen, wie die Beispiele in der ökonomischen Forschung zeigen. Methodologisch bietet die Simulation sich auch zur Prüfung von Modellen an, deren Variation durch Statistiken, z. B. Wirtschaftsstatistiken, extern validierbar ist. Sie erlaubt außerdem, die Gültigkeit des Modells vorausgesetzt, Optimierungsprobleme zu lösen, z. B. bei der Verteilung von Schulplätzen oder der Auswahl von Standorten (vgl. hierzu ABERTSON 1968, GUETZKOW 1962, MEISSNER 1970).

Zu den Abwandlungen der experimentellen Methodik zählen auch **geplante Eingriffe** während einer Studie mit der Methode der Beobachtung, z. B. die Manipulation des Blickkontakts in einem therapeutischen Gespräch. Die enge Beziehung zwischen Beobachtung und Experiment sollte jedoch nicht zu einer Gleichsetzung beider Methoden führen, da es beim Experiment nicht um die Methode der Protokollierung, sondern um die methodologische Anlage der Untersuchung geht.

Ebenso sind geplante und kontrollierte Veränderungen des Feldes während einer teilnehmenden Beobachtung oder Interventionen im Rahmen einer Aktionsforschung Varianten der quasi-experimentellen Verfahren, wenigstens nur selten alle Anforderungen dieser Untersuchungspläne hinsichtlich der Kontrolle der Variablen erfüllbar sind.

Hingewiesen sei schließlich auf **Handlungsrollen**, in denen Personen vorgegebene Positionen mit definierten und limitierten Handlungschancen zugeordnet werden. Das Ziel ist, komplexe soziale Strukturen abzubilden, um die Interaktionen der Personen zu studieren; darüber hinaus sollen sie selbst die Struktur und Zwänge der Situation erkennen können.

5.11.3. Anwendung

Laborexperimente sind bislang eine Domäne der psychologischen und sozialpsychologischen Forschung: Denk- und Wahrnehmungspsychologie, Prozesse interpersoneller Wahrnehmung, Entscheidungstheorie, Aggres-

sionsforschung, lerntheoretische Forschungen oder Probleme der kognitiven Dissonanz sind fast ausschließlich in Laborexperimenten untersucht worden. Dies sind nur einige Beispiele aus der Vielzahl der Anwendungsmöglichkeiten des Laborexperiments. Einen ersten Einblick in die für einen Soziologen wichtigen Experimente bietet der instruktive Sammelband von IRL (1969). Für die (seltenen) soziologischen Laborexperimente ist die partielle Abbildung einer militärischen Organisationsstruktur ein Beispiel (vgl. ZEIDLICH 1961).

Daneben kann man eine Tendenz zu Feldexperimenten beobachten, z. B. in der Aggressionsforschung oder in Studien zur Produktivität in Gruppen oder der Effekte unterschiedlicher Führungsstile in Betrieben (vgl. FRENCH 1953). Gerade Organisationen und Institutionen wie Schule und Betrieb bieten vielfältige Möglichkeiten zu feldexperimenteller Forschung und zugleich die Chance, sie mit politischer Praxis zu verbinden.

Für die einzelnen quasi-experimentellen Verfahren geben, vor allem aus dem Bereich der pädagogischen Forschung, CAMPBELL & STANLEY (1963) zahlreichere Anwendungsbeispiele.

5.11.4. Vorgehen

Das Experiment setzt Hypothesen voraus, deren Variablen präzise definiert sind. Um die genaue Beziehung zwischen unabhängigen Variablen und abhängigen Variablen zu bestimmen, sind die **Operationalisierungen** normalerweise strenger als bei anderen Methoden der Sozialforschung, da sie in Laborexperimenten Teil der Konstruktion der Erhebungssituation sind. Soll die Geschwindigkeit der Aufgabenerfüllung einer Gruppe in Abhängigkeit von der Kommunikationsstruktur untersucht werden, so ist die Art der Aufgabe zu spezifizieren; die Struktur mag in der Trennung der Personen durch Kabinen bestehen, die in unterschiedlicher Weise, entsprechend den einzelnen Versuchsbedingungen, durch Kabel und Mikrophone miteinander verbunden sind, so daß immer nur einige Personen verbalen Kontakt zueinander haben.

Der zweite Schritt, in dem Beispiel bereits angedeutet, bezieht sich auf die **Konstruktion der Experimentalsituation**. Hierzu gehört auch die Anweisung, die der Versuchsleiter den Teilnehmern gibt.

Die Zahl der unabhängigen und abhängigen Variablen und ihre vermutete Beziehung bestimmen die Wahl des Untersuchungsplans (design) und seine statistische Auswertung. Hierfür sei auf die einschlägige Literatur verwiesen, z. B. CAMPBELL & STANLEY (1963), EDWARDS (1971), MITTENBECKER (1964), WINER (1962), ZAMERMANN (1972).

Um die Probleme experimenteller Methodik zumindest anzudeuten, seien vier Untersuchungspläne aufgeführt:

Versuchsplan 1.: M₁ X M₂

Die Effekte nur einer unabhängigen Variablen (X) auf eine abhängige Variable werden untersucht. CAMPBELL & STANLEY (1963, S. 177) rechnen diesen Plan zu den »Vorexperimentellen«, da sich die Differenzen zwischen den Messungen M₁ und M₂ nicht nur als Effekt von X erklären lassen, sondern u. a. auch auf historische Veränderungen zwischen den Zeitpunkten der Messung, Reifungsprozesse der Personen, Effekte der Ersmessung zurückgehen können. Da dieser Versuchsplan auch vielen Panel-Studien zugrunde liegt, gelten entsprechend diese Einwände gegen die Validität der Schlüsse.

Versuchsplan 2.: E : M₁ X M₂ K : M₃ M₄

Dieser »echte« und häufig verwendete Versuchsplan erlaubt, die Effekte der Zeit zu kontrollieren, indem zu gleichen Zeitpunkten wie in der Experimentalgruppe (E) auch Messungen bei einer Kontrollgruppe (K) vorgenommen werden. Hierdurch lassen sich z. B. die unter (1) genannten Fehlerquellen bis auf die Effekte der Ersmessung kontrollieren. Zur Auswertung dient meist der t-Test.

Versuchsplan 3.: Faktorieller Versuchsplan

	Y ₁	Y ₂
X ₁	M ₁	M ₂
X ₂	M ₃	M ₄

In diesem quasi-experimentellen Versuchsplan werden die Effekte zweier unabhängiger Variablen (X, Y) in je zwei Ausprägungen auf die abhängige Variable gleichzeitig untersucht. Die Messungen sollten möglichst zum gleichen Zeitpunkt vorgenommen werden, um Effekte der Zeit zu verhindern. Das mathematisch-statische Modell der Auswertung ist zumeist die Varianzanalyse.

Versuchsplan 4.: M₁ M₂ M₃ M₄ X M₅ M₆ M₇ M₈

Eine abhängige Variable, z. B. die Frequenz des Fernsehens, wird periodisch gemessen sowohl vor der Einführung der unabhängigen Variablen, z. B. dem Kauf eines Farbfernsehers, wie später, um die Effekte der unabhängigen Variablen zu analysieren, im Beispiel die Veränderungen der Fernsehfrequenz. Auch die Analyse von Zeitreihen ist ein quasi-experimentelles Verfahren. Die Effekte der unabhängigen Variablen unterliegen mehreren Fehlerquellen, u. a. der rivalisierenden Erklärung durch historische Veränderungen oder reaktive Effekte aufgrund der wiederholten Messungen. — Die Auswertung erfolgt durch Signifikanz-Tests oder Extrapolation von Trends.

Ein zentrales Element der **Fehlerkontrolle** im Experiment ist die **Aufteilung der Teilnehmer an den Versuchen auf die einzelnen Gruppen und Versuchsbedingungen**. Das Ziel ist, möglichst Gruppen gleicher Merkmalsverteilung zu erhalten, um Gruppenmerkmale nicht mit den gewünschten Effekten der unabhängigen Variablen zu vermengen. Um die Gleichartigkeit der Gruppen zu gewährleisten, bedient man sich zweier Verfahren, die

sich – vereinfacht – mit der Quota- und der Zufallsstichprobe vergleichen lassen:

VP – Zufallsstichprobe
1. Bei der **Parallelisierung** (matching) werden die Teilnehmer des Experiments dertart auf die **Experimental- und Kontrollgruppe** aufgeteilt, daß entweder a) jeder Person in der Experimentalgruppe eine andere, hinsichtlich möglichst vieler Merkmale gleiche Person in der Kontrollgruppe entspricht (ideal: jeweils eineiige Zwillinge), oder b) sich nicht die Individuen, sondern die Verteilung der Merkmale in den beiden Gruppen insgesamt weitgehend gleichen (also die Randsummen der Tabelle mit den Merkmalen der Gruppenmitglieder). Die Parallelisierung setzt voraus, daß die Verteilung der Merkmale in den beiden Gruppen gleich ist, so doch wenigstens die Verteilung der Merkmale in den beiden Gruppen bedeusamen Variablen zu kennen, welche einen Effekt auf das Experiment haben könnten.

VP – Zufallsstichprobe
2. Da dies nur selten der Fall ist, sollte eine zufällige Verteilung der Personen auf die Experimental- und Kontrollgruppe resp. zusätzlich auf die einzelnen Versuchsbedingungen erfolgen. Durch diese **Randomisierung** wird eine maximale Zufallsstreuung der Variablen erreicht, die einen verzerrenden Einfluß auf die Ergebnisse haben könnten. Daher ist nach STRANLEY & CAMRELL (1963, S. 185) die Randomisierung in jedem Fall einer Parallelisierung vorzuziehen.

Probleme des Feldexperimentis: Lassen sich die aufgeführten Elemente des Vorgehens auch auf Feldexperimente übertragen? – Methodologisch ist dies zu bejahen: Auch ein Feldexperiment verlangt, die Hypothesen präzise zu formulieren und die Beziehungen zwischen den Variablen anzugeben. Auch die unterschiedlichen Versuchspläne und zuletzt genannten Kontrollverfahren sind anwendbar.

Vorbehalte sollte man hinsichtlich der Möglichkeit haben, alle unabhängigen Variablen zu kontrollieren. Der Forscher hat a) meist nur Teilkenntnisse aller Variablen des Feldes, er kennt b) auch nicht ihre interaktiven Effekte, z. B. zwischen der Zahl der anwesenden Personen, ihrem sozialen Status und ihrem Alter, schließlich kann er c) nur einen Teil dieser Variablen manipulieren.

Mit der Komplexität der Erhebungssituation steigen die Schwierigkeiten, das Feldexperiment wird zur Feldstudie: Der Forscher kann nur noch eine Selektion von Bedingungen und Personen, nicht aber ihre Manipulation vornehmen. Besonders deutlich wird dies an Feldexperimenten in Gemeinden.

FRENCH (1953, S. 103) hat daher vorgeschlagen, sich in Feldexperimenten auf die Analyse der Haupteffekte der Variablen zu konzentrieren, also zunächst nur die stärksten Effekte der unabhängigen Variablen auf die abhängigen zu berücksichtigen, bevor man die Wechselwirkungen zwischen den unabhängigen Variablen untersucht. Ein weiteres **Problem ist das Messniveau: In der Mehrzahl der Fälle dürfte in Laborexperimenten eine höhere Skalengüte zu erreichen sein, mithin auch eine differenziertere statistische Auswertung.**

Um diese Schwierigkeiten zu verringern, empfiehlt sich folgendes Vorgehen:

Angewandte Feldexperimente

1. Sorgfältige Auswahl der Erhebungssituation, explorative teilnehmende Beobachtung der prospektiven Situation, um Informationen zu erhalten, ob sich die Hypothesen angemessen prüfen und die Bedingungen kontrollieren lassen.
2. «Insulation» (FRENCH 1963, S. 117) der Situation vom Rest des Feldes, wodurch der Einfluß weiterer Variablen verringert werden soll. Ein gutes Beispiel ist die Erhebungssituation in der unten dargestellten Studie.
3. Nicht zu geringe Zahl von Experimenten; Replikation des Experiments in anderen Erhebungssituationen.
4. Versuch, mit Kontrollgruppen und Randomisierung zu arbeiten.
5. Verzicht auf Messungen zu Anfang des Experiments, z. B. durch Interviews mit den Teilnehmern, da sonst nur schwer kontrollierbare reaktive Effekte (Messung – Versuchsbedingung) auftreten.

Abschließend wird das experimentelle Vorgehen an einem Feldexperiment eingehend dargestellt. Dazu ist es erforderlich, zunächst die Forschungssituation, der das Experiment zugehört, zu beschreiben.

In dem angesehenen New Yorker Stadtteil Queens wurde an einem Aprilmorgen des Jahres 1964 die Krankenschwester Kitty Genovese überfallen und durch zahlreiche Messerstiche getötet. Der Vorfall dauerte 35 Minuten, auf die Hilfschreie der Überfallenen geschah nichts. Untersuchungen ergaben, daß 38 Personen den Vorfall wenigstens zum Teil beobachtet oder bemerkt hatten, ohne daß eine von ihnen zu Hilfe gekommen wäre oder die Polizei benachrichtigt hätte.

Es handelt sich hier nicht um einen singulären Fall, sondern um ein eklatantes Beispiel unterlassener Hilfeleistung angesichts mehr oder minder großer Not und physischer Bedrohtheit anderer Bürger. Wie läßt sich diese Passivität der Augenzeugen erklären? Die Dimensionen des Problems kann man durch zahlreiche Fragen bezeichnen: Hat man Furcht, selbst verletzt zu werden? Scheuen die Augenzeugen das Risiko des Engagements? Ist die Vertrautheit mit brutalen Aktionen so groß, daß der Einzelfall untergeht oder nur noch veyeuristisch erlebt wird? Ist die unspezifische Richtung des Hilfschreies ein Appell, für den sich in einer Großstadt niemand verantwortlich fühlt? – Die unterschiedlichen Dimensionen, die zudem nicht unabhängig voneinander sind, deuten die Schwierigkeit einer Erklärung des Phänomens «unterlassene Hilfeleistung» durch Augenzeugen an. Eine exakte Erklärung hätte mit zahlreichen, zum Teil interdependenten Variablen zu arbeiten, die sich nur in einem komplexen Kausalmodell abbilden ließen, für das bislang jedoch die theoretische Vorarbeit fehlt. So ergeben sich unterschiedliche Teilprobleme, die experimentell jeweils Aspekte des Problems und seiner Erklärung untersuchen, z. B. die Hypothese von MURGRAM (1970) über das Verhalten von Großstädtern im Vergleich zu Bewohnern kleiner Orte, wobei MURGRAM den Begriff «Anonymi-

tät» durch «Überbelastung der kognitiven Kapazität» zu präzisieren sucht. Aus der Untersuchung der Bedingungen, unter denen Hilfeleistungen unterbleiben, hat sich in den letzten Jahren eine sozialpsychologische Forschungsrichtung unter der Bezeichnung «bystander effect», übersetzbar als «Augenzeugen-Forschung», entwickelt. Dabei sind sowohl Labor- wie Feldexperimente angewendet worden, je nachdem, welche Dimension untersucht wurde und wie viele Variablen in den Untersuchungsplan eingingen.

So konnten LARANÉ & DARLEY (1970) zeigen, daß mit steigender Zahl der Augenzeugen die Wahrscheinlichkeit einer Hilfeleistung geringer wird. In einem Feldexperiment prüften AITKMAN et al. (1969, zit. nach MURKIN 1970, S. 147) die Hypothese, daß Hilfeleistungen seltener in der Stadt als in kleinen Gemeinden gewährt werden. Die Versuchsleiter klingelten an Haustüren, baten dann, telefonieren zu dürfen, weil sie die Adresse eines Freundes in der Nähe verlegte hätten. Die Bereitschaft, jemandem Zutritt zur Wohnung zu gestatten, um zu telefonieren, war im Extremfall fünfmal größer in einer Kleinstadt als in Manhattan.

Die bisherige Forschung hat gezeigt, daß die Hilfeleistung unter anderem abhängig ist von der Art des Notfalls (z. B. Autopanne, Sturz eines Betrunkenen, epileptischer Anfall), der Direktheit des Kontakts zum Ereignis, dem Geschlecht und der Zahl der Augenzeugen sowie vor allem den Reaktionen der Augenzeugen als einem sozialen Verhaltensmodell für den Einzelnen zur Interpretation der Situation (CLARK III & WORD 1972, LARANÉ & DARLEY 1970). Wenn gleich sich Teile des Problems in Laboratoriumssituationen abbilden lassen, ist doch das Feldexperiment angesichts der Komplexität des Problems und des bisherigen Wissensstandes das angemessenere experimentelle Verfahren. Ein weiteres Beispiel ist das Experiment von PILAVIN & PILAVIN (1972), das etwas ausführlicher diskutiert werden soll. Die Studien der Autoren (vgl. PILAVIN, ROBIN & PILAVIN 1969) sind der Problemdimension «Risiko des Augenzeugen» zuzuordnen.

Die Autoren gingen von mehreren Hypothesen aus, die sie in einem zweistufigen Modell mit folgenden Annahmen verbanden: Die Beobachtung eines Notfalls führt zu physiologischer und emotionaler Erregung, die der Augenzeuge zu reduzieren sucht; dazu hat er vier Möglichkeiten: direkte Hilfe, indirekte Hilfe (Suche nach anderen), Nichtstun, Fortgehen. Hieraus leiten die Autoren drei Hypothesen ab:

- a) Mit zunehmender Erregung des Augenzeugen steigt die Wahrscheinlichkeit, daß er eine Reaktion zeigt.
- b) Hält man den Grad der Erregung konstant, so steigt mit zunehmender Höhe der Kosten für die Unterlassung einer Hilfe die Wahrscheinlichkeit der Hilfe gegenüber dem Fortgehen.
- c) Steigen die Kosten für eine Hilfeleistung, so sinkt die Wahrscheinlichkeit direkter Hilfe, es steigt die Wahrscheinlichkeit indirekter Hilfe oder des Fortgehens» (PILAVIN & PILAVIN 1972, S. 353).

In einem Feldexperiment sollten nun diese Hypothesen geprüft werden, wobei die Art des Notfalls (unabhängige Variable) kontrolliert und die Kosten der Hilfe (abhängige Variable) variiert wurden.

Untersuchungsplan: In einen U-Bahn-Wagen stiegen gleichzeitig vier Mitglieder des Forschungsplans: zwei Beobachter, ein «programmierter» Augenzeuge und ein «Opfer». Der erste Beobachter kam durch die mittlere Tür des Wagens und setzte sich auf einen Platz in der Mitte des Abteils, der zweite Beobachter und der programmierte Augenzeuge kamen durch eine Tür am Ende des Abteils und setzten sich auch dorthin. Als letzter kam durch die mittlere Tür das «Opfer», ein älterer männlicher Student, durchschnittlich aussehend, in den Wagen. Nach Anfahren des Zuges ging das «Opfer» an einem Stock von der Abteilmittelle auf das Abteillende zu, an dem der programmierte Augenzeuge saß, und fiel direkt vor ihm zu Boden.

Variiert wurde nun die Art des Notfalls: In der einen Bedingung fiel das «Opfer» zu Boden (niedrige Kosten der Hilfe), in der zweiten kam zusätzlich etwas Blut aus seinem Mund (hohe Kosten der Hilfe). Die zweite Variation bestand in der Art des programmierten Augenzeugen: Er trug entweder die Kleidung eines Pfarrers (Annahme: Rolle des veralgemeinerten Helfers) oder eines Arztes (Annahme: Rolle des spezialisierten Helfers) oder einen normalen Straßenganzug. Es handelt sich um eine Erhebungssituation mit insgesamt sechs verschiedenen Bedingungen, gemessen wurde jeweils die Zeit bis zur Hilfeleistung durch die Augenzeugen, die Zahl der Helfenden und die Art der Hilfe. Kann kein Augenzeuge dem «Opfer» zu Hilfe bemüht werden, so wird die Zeit bis zur Hilfeleistung durch ihn; er geleitete ihn in der nächsten Station aus dem Wagen, wo auch die beiden Beobachter ausstiegen.

Die Feldexperimente wurden alle an einem Tag durchgeführt. Um nicht zu volle Züge zu haben, wurde die Zeit zwischen 15.30–16.30 Uhr und 17.30–18.30 Uhr gewählt. Die Fahrzeit der U-Bahn zwischen zwei Stationen betrug etwa drei Minuten. Kontrolliert wurde auch die Fahrtrichtung des Zuges und die Reihenfolge der beiden Bedingungen im Verhalten des Opfers.

Es traten eine Reihe nicht-antizipierter Konsequenzen in der Erhebungssituation auf: Verwicklungen mit der Polizei, Versuche der Augenzeugen, die Notbremse zu ziehen, die Gefahr einer Panik während einiger Versuche in der «Blut»-Bedingung. Zwei Versuche wurden völlig ausgeschlossen, da eine Schwester in Tracht anwesend war. Diese Vorfälle zeigen nicht nur die experimentellen Schwierigkeiten, sondern auch die ethischen Probleme des Feldexperiments.

Von den insgesamt 40 Versuchen entfielen je die Hälfte auf die Bedingung «Kein Bluten des Opfers» und die Bedingung «Bluten des Opfers». Der programmierte Augenzeuge trug in 13 Fällen die Kleidung eines Opfers, in 12 die eines Pfarrers, in den restlichen 15 die eines normalen Passanten.

Die Ergebnisse in Tabelle 14 belegen, daß unter der «Blut»-Bedingung seltener Hilfe geleistet wurde und die Zahl der Helfer niedriger war als in der anderen Bedingung. Darüber hinaus bestand ein signifikanter Unterschied zwischen den beiden Bedingungen hinsichtlich der Zeit bis zur Hilfeleistung: vom Fall des «Opfers» bis zur Hilfe vergingen im Median (Punkt

der Prozentverteilung, vor dem 50 Prozent aller Fälle liegen) in allen Bedingungen 12 Sekunden, nicht einbezogen indirekte Hilfe, wie das Angebot, Hilfe zu holen. Der Median der Reaktionszeit in der «Blut»-Bedingung lag bei 21,5 Sekunden, in der Bedingung ohne Blut bei nur 8,5 Sekunden. Nimmt man diese Ergebnisse zusammen, so hat sich die Hypothese bewährt, daß bei hohen Kosten der Hilfe die Hilfe seltener und langsamer erfolgt als in der weniger dramatischen Bedingung mit niedrigen Kosten.

Tabelle 14: Art der Hilfe und Zahl der direkt Helfenden nach Versuchsbedingungen (PILAVIN & PILAVIN 1972, S. 357)

Bedingung	keine Hilfe		nur indirekte Hilfe		Zahl der direkt Helfenden			Summe
	4	1	3	0	1	2	3 +	
«Opfer» blutend	4	1	3	0	3	4	6	20
«Opfer» nicht blutend	1	1	6	12	1	2	1	20

«Direkte Hilfe» kontra «indirekte Hilfe» + «keine Hilfe»: Fisher's Exact Probability Test (zweiseitige Fragestellung) $p = .01$. Für Zahl der direkt Helfenden (0, 1, 2, 3 +) $\chi^2 = 13,1$, $df = 3$, $p < .01$.

Tabelle 15: Median der Reaktionszeiten nach Versuchsbedingungen und Art des programmierten Augenzeugen (PILAVIN & PILAVIN 1972, S. 358)

Bedingung	in Sekunden *				Summe
	Art des programmierten Augenzeugen	Arzt	Pfarrer	keiner	
«Opfer» blutend	n	60,00	20,50	18,00	21,50
	n	7	6	7	20
«Opfer» nicht blutend	n	6,00	11,75	10,75	8,50
	n	6	6	8	20

* keine Reaktion wurde als Reaktionszeit von 60 Sekunden berechnet.

Wie sich die Art des programmierten Augenzeugen auf das Verantwortungsbewußtsein der anderen Augenzeugen auswirkt, zeigt Tabelle 15. Obgleich hier Unterschiede in der Geschwindigkeit der Reaktion bestehen, sind diese zwischen den einzelnen programmierten Augenzeugen-Bedingungen nicht signifikant. Eine zweite Varianzanalyse der Ergebnisse erbrachte, daß die einzig signifikante Quelle der Varianz der Ergebnisse die Variable «Blut» – «kein Blut» war. Hingegen erwiesen sich weder die Art des programmierten Augenzeugen noch interaktive Effekte zwischen Art des Augenzeugen und den «Blut» – «kein Blut»-Bedingungen als

statistisch bedeutsam. Dennoch ist den Autoren zuzustimmen, wenn sie von einer «Diffusion der Verantwortung» (DARLEY & LATANÉ 1968) in der Versuchsbedingung «Blut» – «Arzt» sprechen, d. h. die Zuschauer entledigten sich der eigenen Verantwortung, weil eine Person anwesend war, von der man erwarten konnte, daß sie ihrer Verpflichtung zur Hilfe nachkommt.

Über diese Ergebnisse hinaus belegte die Studie eine signifikant höhere Zahl von Männern unter den Augenzeugen, die als erste zu Hilfe kamen; nicht bestätigen konnte sie die Hypothese, daß mit steigender Zahl von Augenzeugen auch die Hilfsbereitschaft sinkt.

Die relativ ausführliche Darstellung des Experiments veranschaulicht, welche Schwierigkeiten auftreten, wenn Hypothesen in einem strengen experimentellen Verfahren überprüft werden sollen. Dem Experiment liegt ein vergleichsweise einfacher Untersuchungsplan zugrunde (Nr. 4 unter den oben genannten). Wenige Variablen mit nur wenigen Ausprägungen wurden in einer entsprechend geringen Zahl von experimentellen Bedingungen untersucht. Diese und die anderen erwähnten Experimente zum Augenzeugen-Effekt bilden jeweils nur Teile des Problembereichs ab, bedürfen dazu aber z. T. bereits recht komplizierter Erhebungssituationen. Ein Vorreil des Experiments von PILAVIN & PILAVIN ist, daß die abhängige Variable auf einer Rattoskala gemessen werden konnte: Hilfsbereitschaft in Sekunden der Reaktion.

An den zitierten Feldexperimenten ist erkennbar, daß es sehr wohl möglich ist, alltägliche Situationen als Erhebungssituation zu nutzen, um in ihnen soziale Prozesse experimentell zu studieren. Dazu bedarf es sicherlich neben der theoretisch angemessenen Auswahl der Situationen auch einiger sozialer Fantasie. Schließlich sei nochmals betont, daß die Experimente jeweils Teile von Forschungsrichtungen und Teilprüfungen von Hypothesen darstellen, daher notwendig einer schrittweisen und kumulativen Forschung bedürfen, um nicht als Einzelbefunde ohne Bezug zu einer generalen Theorie zu bleiben.

5.11.5. Erhebungssituation

In der Erhebungs-, oder hier genauer: der Experimentalsituation, treten vor allem zwei Probleme auf: die reaktiven Effekte des Forschers auf die am Experiment beteiligten Personen und die reaktiven Effekte der Teilnehmer. Die Probleme sind denen im Interview ähnlich.

Einer der reaktiven Effekte ist die Erwartung des Forschers auf das Ergebnis der Untersuchung, der «experimenter bias», der sich in Studien mit allen Methoden nachweisen läßt. Von den zahlreichen Experimenten, die ROSENTHAL (1966) durchgeführt hat, sei nur eines angeführt: Einer Gruppe von Versuchsteilern, die Lernexperimente an Ratten vornehmen sollten,

wurde gesagt, die Ratten, die sie erhalten hätten, seien durch Züchtung besonders klug, einer anderen Gruppe von Versuchsleitern wurde gesagt, ihre Ratten seien durch Züchtung besonders dumm. Obgleich kein Unterschied zwischen den beiden Rattenarten bestand, erzielten bei gleicher Versuchsanordnung die ersten Versuchsleiter signifikant bessere Lernerfolge bei ihren Ratten als die zweiten (ROSENTHAL 1963). Andere Experimente haben nachgewiesen, daß die Auswahl der Versuchsleiter nach Schicht, ethnischer Zugehörigkeit, Alter und Persönlichkeitsmerkmalen (z. B. Ängstlichkeit) von Einfluß auf die Ergebnisse des Experiments sind. Der Versuchsleiter ist selbst ein «Stimulusobjekt» (MCGURCAN 1963). Weiterhin sind das nonverbale Verhalten des Versuchsleiters und die Art seiner Anweisungen bedeutsam. Um die Effekte des Versuchsleiters zu kontrollieren, bieten sich u. a. veränderte Versuchspläne oder die Einsetzung eines Hilfsversuchsleiters an, der nicht über die Hypothesen des Experiments informiert ist.

Die Teilnehmer am Laborexperiment nehmen eine Interpretation der Situation vor. Sehen sie die Situation als real an, dann kann man auch schwierig von einer Künstlichkeit des Experiments sprechen (OPR 1970 a). Teil dieser Interpretationen sind Überlegungen über den Zweck des Experiments. Die von den Teilnehmern vermutete Hypothese oder das Bemühen, sich an vermutete Erwartungen des Versuchsleiters anzupassen, ja überhaupt an dem Experiment teilzunehmen, sind unter der Bezeichnung «demand characteristics» zusammengefaßt worden. Zur Kontrolle dieser Effekte bietet sich u. a. an, die Teilnehmer über den wahren Zweck des Experiments durch Äußerung einer falschen Hypothese zu täuschen oder sie auf die Wissenschaftlichkeit des Experiments zu verpflichten (vgl. hierzu eingehend ZIMMERMANN 1972, Kapitel 13).

5.11.6. Stichprobe

Unter Stichprobe soll nicht verstanden werden, wie die Teilnehmer durch Randomisierung den einzelnen Versuchsbedingungen zuzuordnen sind, sondern die **Auswahl der am Experiment beteiligten Personen**. In der Literatur zum Experiment ist häufig von einer «Zufallsauswahl der Versuchspersonen» die Rede. Eine Analyse der Stichproben, die tatsächlich in Laborexperimenten verwendet werden, zeigt jedoch nachgerade das Gegenteil. Wie am Beispiel im Abschnitt 3.4 belegt, **handelt es sich selten um Zufallsauswahlen aus der Gesamtbevölkerung, sondern um bewußte Auswahlen aus einer Subpopulation**; zumeist Studenten, oder gar um Personen, die sich freiwillig zur Teilnahme am Experiment gemeldet haben. Sofern man nicht unterstellt, die hier in Laborexperimenten an solchen Stichproben geprüften Hypothesen enthielten nur Variablen, die zum Beispiel schicht- und altersinvariant seien, fällt ein großer Teil soziologisch relevanter Va-

riablen und situativer Effekte in die Gruppe der nicht-manipulierten unabhängigen Variablen. Da die genannte Annahme wohl kaum gerechtfertigt ist, bestehen erhebliche Zweifel an der Generalisierbarkeit der Ergebnisse auf andere als die jeweils untersuchten Gruppen, zumal Replikationen von Experimenten mit anderen sozialen Gruppen selten sind. Eine dieser Ausnahmen sind die Experimente von MINGRAM (1966) – was immer man sonst gegen sein Auswahlverfahren einwenden mag.

Warum die Forderung, exaktere Auswahlpläne zu verwenden, so schwer erfüllbar ist, geht aus den Voraussetzungen des Experiments hervor: 1. Es bedürfte selbst bei sukzessiv replizierenden Experimenten eines sehr aufwendigen Untersuchungsplans, um auch nur die wichtigsten demographischen Variablen neben den anderen unabhängigen Variablen zu kontrollieren. 2. Experimente sind aus ethischen Gründen auf jene Freiwilligkeit der Mitarbeit angewiesen, die auch alle anderen sozialwissenschaftlichen Methoden kennzeichnet. Stichprobentheoretischen Anforderungen angemessener experimentelle Verfahren dürfen eher in den Formen des Feldexperimentes oder den quasi-experimentellen Untersuchungen mit Längsschnitt-Analysen zu sehen sein.

5.11.7. Pretest

Es ist fraglich, ob man bei experimentellen Verfahren überhaupt mit Pretests arbeiten kann, weil z. B. kein Instrument zu testen ist. **Ein Pretest wird sich deshalb am ehesten auf die Erhebungssituation beziehen: auf die Künstlichkeit und die Reaktionen der Personen im Laborexperiment oder die Abgrenzbarkeit und Störinflüsse beim Feldexperiment.** (Auch in der zitierten Untersuchung über Hilfeleistungen von Augenzeugen wurde ein Pretest durchgeführt.) Schließlich ist bei Feldexperimenten zu prüfen, ob «die manipulierte Variable überhaupt einen Effekt auf die abhängige Variable hat» (ZIMMERMANN 1972, S. 196).

Übersicht 39 stellt nur einen Versuch dar, angesichts der Vielzahl experimenteller Untersuchungspläne einige Kriterien für einen Pretest zu entwickeln.

Übersicht 39: Aufgaben des Pretests bei einem Experiment

Umfang:	Für Laborexperimente ein Pretest, für Feldexperimente zwei bis drei Pretests.
Legitimation:	Erläuterung des Untersuchungsziels vor – soweit möglich – und nach dem Experiment. Haben die Teilnehmer ein- gewilligt?
Erhebungssituation:	Zeitpunkt und Ort richtig gewählt? Verfassung der Personen? Situation vor dem Experiment? Verfassung und

Situation nach dem Experiment? Reaktive Effekte des Forschers? Täuschung der Teilnehmer erforderlich? Realitätsgehalt des Laborexperiments für Teilnehmer? Läßt sich der geplante Ablauf durchführen? Sind die »Stimuli« eindeutig? Störfaktoren während des Experiments? Nicht-antizipierte Konsequenzen? Hinreichende Unterschiede zwischen Experimental- und Kontrollgruppe? Kontrolle der Zuweisung durch Randomisierung möglich? Effekte der unabhängigen Variablen auf abhängige vorhanden? Sind Experimental- und Kontrollgruppe dem Forschungsplan angemessen zusammengesetzt? Auf welche Grundgesamtheit kann generalisiert werden (Personen, Situationen, Verhaltensformen u. a.)? Moralität der Stichprobe bei Messungen zu mehreren Zeitpunkten? Konsequenzen der Selbst-Selektivität der Stichprobe?

Kontrollen:
Befragung vor und nach dem Experiment. Aufzeichnung des Experiments durch Kamera oder Beobachter.

5.11.8. Fehlerquellen

Die Frage nach der Angemessenheit der Stichprobe und die Diskussion über die Künstlichkeit von Experimental-Situationen beziehen sich beide auf die Generalisierbarkeit der Ergebnisse. Wenn nur bestimmte Personen an Experimenten teilnehmen – für welche Gruppe der Bevölkerung stehen sie ein? Wenn das Experiment in einer weitestgehend kontrollierten Situation stattfindet – auf welche Situationen oder Variablen-Strukturen außerhalb des Labors sind die Ergebnisse übertragbar?

Um die Gültigkeit von Experimenten näher zu bestimmen, haben CAMPELL & STANLEY (1963) die grundlegende Unterscheidung zwischen interner und externer Validität eingeführt. Die **interne Validität** bezieht sich auf die **Eindeutigkeit der Messung im Experiment**, ob also in den Untersuchungsplan oder die Erhebungssituation nicht weitere unkontrollierte Variablen eingegangen sind, die mit der Messung der Effekte der unabhängigen Variablen auf die abhängige Variable vermengt werden können. Eine ausführliche Darstellung der Effekte auf die interne Validität geben CAMPELL & STANLEY; einige dieser Effekte wie historischer Wandel, Reifung oder Meßeffekte wurden bereits im Abschnitt 5.11.4 angeführt. Der Kontrolle dieser Fehler dienen u. a. die Randomisierung der Teilnehmer auf die einzelnen experimentellen Bedingungen. Die interne Validität der einzelnen Versuchspläne ist unterschiedlich gut, worauf ebenfalls im Abschnitt 5.11.4 hingewiesen wurde. Hiervon folgt auch, daß die interne Validität bei Feldexperimenten im allgemeinen niedriger und schwerer zu messen ist.

Die interne Validität ist eine notwendige Bedingung der **externen Validität**. Letztere bezieht sich auf die **Generalisierbarkeit der Ergebnisse**. Als

Verringerung der externen Validität können sich z. B. die reaktiven und interaktiven Effekte der Messung auswirken. So ist die externe Validität der aufgeführten Versuchspläne gering, sofern nicht zusätzliche Kriterien eingeführt werden. Im allgemeinen steigt die externe Validität der Laborexperimente mit der Zahl der Replikationen, der Repräsentativität der Versuchsteilnehmer und der Nicht-Falsifikation der Hypothese bei unterschiedlichen Operationalisierungen (obgleich das wiederum zu der im Abschnitt 2.3 diskutierten Frage nach der Stichprobe aus einem Universum aller Items/Indikatoren führt). Aufgrund der Realitätsnähe wird dem Feldexperiment eine höhere externe Validität zugesprochen.

Eine ausführliche Darstellung der Validitätsprobleme experimenteller Methoden geben CAMPELL & STANLEY (1963); verwiesen sei auch auf BREDENKAMP (1969, S. 354 ff.) und ZIMMERMANN (1972).

5.12. SEKUNDÄRANALYSE

5.12.1. Voraussetzungen

Die Sekundäranalyse ist eine Methode, bereits vorhandenes Material (Primärerhebung) unabhängig von dem ursprünglichen Zweck und Bezugsrahmen der Datensammlung auszuwerten.

Der Terminus wurde 1950 durch KENDALL & LAZARFIELD (1950) eingeführt, um das methodische Vorgehen zu kennzeichnen, das STROUFER und Mitarbeiter in ihrer Studie »The American Soldier« anwendeten, als sie 300 Umfragen des War Department, Information and Education Division, erneut auswerteten. Andere Autoren haben die Termini »Re-Analyse« (z. B. KENDALL & LAZARFIELD 1950, RUEY 1963) oder »erweiterte Analyse« (HARDER 1969) vorgeschlagen; doch hat sich der Begriff Sekundäranalyse durchgesetzt. Die Methode ist sehr viel älter als der Begriff: ENGELS ver-wendete in seinem Bericht über »Die Lage der arbeitenden Klasse in England« aus dem Jahre 1845 u. a. häufiger Statistiken der vom britischen Innenministerium eingesetzten Poor Law Commission. DURKHEIM hat in seiner 1897 erschienenen Studie über Erscheinungsformen und Ursachen des Selbstmordes zusammen mit MAUSS über 26 000 Fälle analysiert.

Zwei Arten von Sekundäranalysen lassen sich aufgrund des Primärmaterials unterscheiden:

1. Sekundäranalysen von sozialwissenschaftlichen Erhebungen und Meinungsumfragen. Zu den Studien liegen auch Publikationen vor, in denen das Material interpretiert wurde.
2. Sekundäranalysen von demographisch-statistischem Material, z. B. amtlichen Statistiken: Volkszählung, Gebäude- und Wohnungszählung, Statistiken der Meldeämter, der Polizei, der Gesundheitsbehörden. Oder nicht-amtlichen Statistiken: Ortskrankenkasse, Versicherungen, Automobilclubs, Krankenhäuser, Betriebsstatistiken.