

Helmut Kromrey
**Empirische
Sozialforschung**
11. Auflage

Lucius
& Lucius **UTB**

Empirische Sozialforschung

Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung

Von Helmut Kromrey

11., überarbeitete Auflage

2006. 565 S., kt. € 14,90 / sFr 26,80
UTB 1040. ISBN 3-8252-1040-5

Die in elfter Auflage vorliegende Einführung in die standardisierte empirische Sozialforschung ist anwendungspraktisch orientiert. Ihr Aufbau orientiert sich am Ablauf eines realen Forschungsprozesses. Die grundlegenden wissenschaftstheoretischen und methodologischen Fragen werden nicht als Selbstzweck, sondern entsprechend ihrer Relevanz für die Forschungspraxis abgehandelt und durch sozialwissenschaftliche Beispiele veranschaulicht. Der Text setzt Vorkenntnisse nicht voraus und ist besonders an den Bedürfnissen der "Neueinsteiger" orientiert. Für "Wiederholer" und Praktiker werden in den Fußnoten fortlaufend Hinweise zur Vertiefung geboten sowie Querbezüge zu qualitativen Forschungsansätzen hergestellt.

Inhaltsübersicht

- 1 Empirische Sozialforschung und empirische Theorie**
 - 1.1 Vorbemerkungen zur Situation empirischer Sozialwissenschaft
 - 1.2 Grundpositionen der Erfahrungswissenschaft
 - 1.3 Empirische Sozialforschung als "kritisch-rationale Wissenschaft"
 - 1.4 Empirische Verfahren und alternative Wissenschaftspositionen
- 2 Forschungsfragen, Forschungsdesign, Forschungsprozess**
 - 2.1 Fragestellungen und Forschungsansätze: Einige Beispiele
 - 2.2 Der Forschungsprozess als eine Reihe ineinander verzahnter Entscheidungen
 - 2.3 Entdeckungs-, Begründungs-, Verwertungszusammenhang: Das Problem der Wertungen in der empirischen Forschung
 - 2.4 Forschungsplanung und Forschungsdesign
- 3 Die empirische "Übersetzung" des Forschungsproblems**
 - 3.1 Problempräzisierung und Strukturierung des Untersuchungsgegenstandes: dimensionale und semantische Analyse
 - 3.2 Beispiel einer dimensionalen Analyse: Berufserfolg und soziale Herkunft
 - 3.3 Beispiel einer semantischen Analyse: der theoretische Begriff "Entfremdung"
 - 3.4 Zusammenfassung: semantische und dimensionale Analyse im Vergleich
 - 3.5 Begriffe und Definitionen
- 4 Strategien der Operationalisierung und Indikatorenauswahl**
 - 4.1 Indikatoren
 - 4.2 Indexbildung
 - 4.3 Operationalisierung
- 5 Messung und Datenerhebung in den Sozialwissenschaften**
 - 5.1 Die Informationsgewinnung im Prozess der empirischen Forschung
 - 5.2 Exkurs: Die Rolle der Statistik bei empirischen Untersuchungen
 - 5.3 Variablenbildung – Messen – Datenmatrix
 - 5.4 Vertiefung: Die axiomatische Mess-theorie
 - 5.5 Messen durch Indizes (Indexmessung)
 - 5.6 Der sozialwissenschaftliche Spezialfall: Messen durch Befragung
 - 5.7 Zuverlässigkeit (Reliabilität) der Messung
- 6 Auswahlverfahren**
 - 6.1 Zentrale Begriffe: Grundgesamtheit, Auswahl-, Erhebungs- und Untersuchungseinheiten
 - 6.2 Anforderungen an die Stichprobenkonstruktion
 - 6.3 Typen von Auswahlverfahren (Überblick)
 - 6.4 Nicht zufallsgesteuerte Auswahlverfahren
 - 6.5 Zufallsgesteuerte Auswahlverfahren
 - 6.6 Zusammenfassung: Vor- und Nachteile der verschiedenen Auswahlverfahren
- 7 Datenerhebungsverfahren und -instrumente der empirischen Sozialforschung**
 - 7.1 Empirische Inhaltsanalyse
 - 7.2 Beobachtung
 - 7.3 Befragung
 - 7.4 Vergleich der Erhebungsverfahren Inhaltsanalyse, Beobachtung, Befragung
- 8 Methoden und Modelle der deskriptiven Statistik**
 - 8.1 Einige zentrale Begriffe
 - 8.2 Univariate Statistik
 - 8.3 Bivariate Statistik
- 9 Typen und Konzepte empirischer Sozialforschung: Eine Übersicht**
 - 9.1 Spezielle Untersuchungsanordnungen
 - 9.2 Alternative Forschungsparadigmen: Qualitative Sozialforschung und Aktionsforschung

Bestellen Sie per Briefpost, Fax oder eMail
Fax 0711/780 13 76 utb@utb-stuttgart.de



UTB FÜR WISSENSCHAFT · Stuttgart
Tel. 0711/782 95 55-0 · FAX 0711/780 13 76
utb@utb-stuttgart.de · www.utb.de
Lucius & Lucius Verlagsges. mbH · Stuttgart
Tel. 0711/24 20 60 · Fax 0711/24 20 88
lucius@luciusverlag.com · www.luciusverlag.com



(Friedrichs 1977, 54). Jede Untersuchung erweitert Wissen über soziale Zusammenhänge, kann zur Lösung sozialer Probleme beitragen (wenn die neu gewonnenen Kenntnisse mit diesem Ziel eingesetzt werden) oder diese Lösung verhindern bzw. zumindest verzögern (wenn Kenntnisse bewusst zurückgehalten werden, oder wenn durch die Untersuchung bestimmter sozialer Fragen andere Probleme unbeachtet bleiben).

2.4 Forschungsplanung und Forschungsdesign

An mehreren Stellen wurde bereits darauf hingewiesen, dass die Methodenlehre der empirischen Sozialforschung kein Handbuch mit Rezepten bereitstellen kann, die schematisch auf konkrete Forschungsprojekte anzuwenden wären. Vielmehr muss in der Regel – mit Ausnahme weniger, häufig wiederholter Routineerhebungen – für jedes Vorhaben ein zur jeweiligen Thematik „maßgeschneiderter“ Forschungsplan entworfen werden, müssen die jeweiligen Forschungsfragen operationalisiert, müssen problemadäquate Erhebungsinstrumente entwickelt und getestet werden. Dabei können die verschiedenen Verfahren und Techniken aus dem Baukasten der Methodenlehre in vielfältiger Weise miteinander kombiniert werden, so wie sie bei gegebenen Rahmenbedingungen (Zeitrestriktionen, Ressourcen, Stand der Vorkenntnisse, Zugänglichkeit und Kontrolle des Untersuchungsfeldes, Art der verfügbaren Informationsquellen u.a.) für den jeweiligen Untersuchungszweck am nützlichsten zu sein scheinen. Solche Entwurfsarbeit ist vergleichbar der Tätigkeit eines Architekten, der ein Gebäude zu planen hat, das in bestmöglicher Weise der vorgesehenen künftigen Nutzung gerecht werden soll, für dessen Realisierung der Architekt jedoch den vorgegebenen Möglichkeitsrahmen (verfügbare Finanzen, Art und Lage des Grundstücks, örtliche baurechtliche Auflagen usw.) nicht durchbrechen kann.

Trotz aller „Individualität“ des jeweiligen konkreten Forschungsvorhabens gibt es dennoch (in etwas gröberer Perspektive betrachtet) für bestimmte Gruppen von Fragestellungen durchaus einige spezifische Designtypen mit jeweils spezifischer Forschungslogik – wie ja auch der Architekt unterschiedliche Gebäudetypen kennt: vom Einfamilienreihenhaus über den freistehenden Bungalow bis zu Wohn- bzw. Bürohochhaus oder Fabrikhalle.

Aus dem Spektrum der existierenden Designtypen sollen im Folgenden vier besonders häufig vorkommende herausgegriffen und etwas näher dargestellt werden:

- die theorie- oder hypothesentestende Untersuchung,
- das Experiment sowie quasi-experimentelle Ansätze,
- das Standardmodell der Programm-Evaluation sowie
- das deskriptive Surveymodell (Querschnitterhebung nichtexperimenteller Daten).

Zum besseren Vergleich der ihnen zugrunde liegenden Verfahrenslogik wird das von *Hempel* und *Oppenheim* vorgeschlagene Schema deduktiv-nomologischer wissenschaftlicher Erklärung (*H-O-Schema*) herangezogen, das daher zunächst in seinen Grundzügen erläutert werden soll.

2.4.1 Das Modell wissenschaftlicher Erklärung von Hempel und Oppenheim

Der Ausgangspunkt: Es ist etwas – möglicherweise Unerwartetes – passiert, das erklärt werden soll. In der Sprache der Methodologen: Ein aufgetretenes singuläres Ereignis soll logisch haltbar mit empirischen Argumenten erklärt werden. Nochmals anders ausgedrückt: Für einen „zu erklärenden Sachverhalt“ (Explanandum) wird ein „erklärendes Argument“ (Explanans) gesucht.

Wird – wie im analytisch-nomologischen Wissenschaftsverständnis – Kausalität als Wirkungsprinzip in der empirischen Realität unterstellt (vgl. Abschnitt 1.2.2), dann kann der „zu erklärende Sachverhalt“ nur als Effekt einer wirksam gewordenen Ursache (oder mehrerer Ursachen) verstanden werden. Die „Erklärung“ bestünde also in der Angabe der Ursache(n).

Nehmen wir als Beispiel für einen „erklärungsbedürftigen“ Sachverhalt: „Die Studiendauer der Studierenden an deutschen Universitäten ist lang“ (nach Auffassung mancher Hochschulpolitiker: zu lang). Als Erklärung dafür wird von manchen angeboten (z.B. vom SPIEGEL): „Die Lehre an deutschen Universitäten ist schlecht“ (1993, Nr. 16). Als alltägliche Erklärung mag das bereits ausreichend und einleuchtend sein. Als logisches Argument allerdings ist diese „Erklärung“ unvollständig. Es fehlt noch die Angabe eines Ursache-Wirkungs-Prinzips. Dieses könnte in unserem Beispiel etwa lauten: „Wenn die Lehre an der Hochschule

schlecht ist, dann ist die bis zum erfolgreichen Studienabschluss erforderliche Studiendauer lang.“

Zusammengefasst:

<i>Explanans:</i> („das Erklärende“)	(1) Wenn die Lehre an der Hochschule schlecht ist, dann ist die bis zum erfolgreichen Studienabschluss erforderliche Studiendauer lang.	<i>(Ursache-Wirkungs-Prinzip)</i>
	(2) Die Lehre an deutschen Universitäten ist schlecht.	<i>(Vorliegende Ursache)</i> ¹¹
<i>Explanandum:</i> („das zu Erklärende“)	(3) Die Studiendauer an deutschen Universitäten ist lang.	<i>(Eingetretene Wirkung)</i>

Damit aus dieser „Alltagserklärung“ eine wissenschaftliche Erklärung wird, müssen allerdings einige ergänzende Bedingungen erfüllt sein: Zum einen muss es sich bei (1) um ein „empirisches Gesetz“ handeln, des weiteren müssen die in den Aussagen verwendeten Begriffe präzise definiert und operationalisiert sein,¹² und selbstverständlich müssen die Aussagen (2) und (3) empirisch „wahr“ sein.

Formaler ausgedrückt lauten diese „*Adäquatheitsbedingungen für Erklärungen singulärer Ereignisse*“ nach Opp (1976, 128 f. und 131 ff.):

- 1) Das Explanandum muss aus dem Explanans korrekt gefolgert worden sein. Dies bedeutet:
 - a) Die Anfangsbedingungen (s. Fußnote 11) stellen raum-zeitlich lokalisierte Elemente derjenigen Objekte dar, die die Wenn-Komponente bezeichnet.
 - b) Die Prädikate der Dann-Komponente und des Explanandums müssen entweder identisch sein, oder das Prädikat der Dann-Komponente muss in dem Prädikat des Explanandums enthalten sein.
 - 2) Das Explanans muss mindestens ein Gesetz enthalten, das für die Ableitung des Explanandums erforderlich ist, sowie singuläre Sätze, die die Randbedingungen beschreiben.
 - 3) Das Explanans muss empirischen Gehalt haben, d.h. es muss empirisch überprüfbar sein.
- 11 Die in Aussage (2) beschriebenen Situationskomponenten, also die in der Argumentation als „Ursache“ genannten Sachverhalte, werden im Erklärungsschema auch als „Anfangsbedingungen“ oder „Randbedingungen“ bezeichnet.
- 12 Das heißt: Die ‚Qualität der Lehre‘ sowie die ‚Studiendauer‘ müssen gültig und zuverlässig gemessen werden können; und es ist anzugeben, ab welchem Messwert die Lehre „schlecht“, ab welchem Messwert die Studiendauer „lang“ ist.

- 4) Die Sätze des Explanans müssen wahr sein (oder zumindest – weniger anspruchsvoll - : sie müssen sich sehr gut bewährt haben).

Ganz ähnlich ließe sich eine Beobachtung wie die folgende erklären: „Die Studiendauer an deutschen Universitäten ist heute länger als sie vor 20 Jahren war.“ Zutreffen müssten ein Ursache-Wirkungs-Prinzip wie „Je schlechter die Lehre ist, desto länger ist die bis zum Studienabschluss benötigte Zeit“ sowie die empirische Feststellung der „Randbedingung“: „Die Lehre an deutschen Universitäten ist heute schlechter als vor 20 Jahren“.

Allgemein lautet das deduktiv-nomologische Erklärungskonzept von *Hempel* und *Oppenheim* (DN-Schema oder H-O-Schema):¹³

- Explanans:* (1) Es gilt (mindestens) ein nomologisches Gesetz (z.B.: „Wenn A und B, dann C“)
 (2) Die in der Wenn-Komponente genannten Randbedingungen sind erfüllt (z.B.: „A und B liegen vor“)
- Explanandum:* (3) Singulärer Satz, der den zu erklärenden Sachverhalt beschreibt (z.B. „C liegt vor“).

Gegeben ist das zu erklärende „singuläre Ereignis“ (3), *gesucht* ist das „Explanans“ (1 und 2). Bei dieser Art von Erklärung muss (3) deduktiv-logisch aus (1) und (2) folgen, wobei (2) aus der Wenn-Komponente und (3) aus der Dann-Komponente des nomologischen Gesetzes abgeleitet wird.¹⁴

Äußerlich ähnlich ist die Struktur bei den Argumentations-Typen „Prognose“ und „Technologische Aussage“. Bei der „Prognose“ ist das Explanandum aus Gesetz(en) und Randbedingung(en) „vorherzusagen“. Bei „Technologischen Aussagen“ ist das Explanandum das in der Zukunft zu erreichende Ziel; gesucht sind geeignete Gesetze, die in der Dann-Komponente das Explanandum enthalten und die in der Wenn-Komponente Sachverhalte benennen, welche gezielt so veränderbar sind, dass dadurch das gewünschte Ziel erreicht wird.

13 *Hempel*, Carl G.; *Oppenheim*, P. C., 1948: Studies in the Logic of Explanation; in: Philosophy and Science, Vol. 15, 135-175.

14 Analog ist der Argumentationszusammenhang bei einer Je-desto-Beziehung. Das „Gesetz“ (1) könnte lauten: „Je höher X, desto höher Y“. Empirisch erfüllt sein müsste die Randbedingung (2): „X ist gestiegen“. Erklärt wäre damit das Explanandum (3): „Y ist gestiegen“.

2.4.2 Das Design hypothesentestender und theoriebildender Forschung

Vielen erscheint der Theorie- bzw. Hypothesentest als der Prototyp empirischer Forschung: „Sozialwissenschaftliche Datenerhebung ist kein Selbstzweck. Befragungen, Beobachtungen und andere Erhebungsmethoden dienen in der Regel einem allgemeineren Zweck: der Überprüfung der Geltung von Theorien“ (Esser 1984, I, 3). Methodologische Texte sind im Allgemeinen (explizit oder zumindest implizit) vor diesem Hintergrund geschrieben, methodologische Argumentationen und Dispute häufig nur vor diesem Hintergrund zu verstehen. Auch Forschungsberichte über abgeschlossene Projekte berufen sich häufig auf die Logik des hypothesentestenden Vorgehens.¹⁵

Die *Fragestellung* dieses Typs von Forschung ist: Kann eine Hypothese oder eine Theorie empirische Geltung beanspruchen? Das mit der Beantwortung dieser Frage verfolgte *Ziel* könnte sein:

- eine bei der Analyse empirischer Befunde einer früheren Untersuchung ad hoc formulierte Dateninterpretation auf ihre allgemeinere Geltung zu testen oder
- zwischen „konkurrierenden“ Hypothesen eine empirische Entscheidung herbeizuführen oder
- Theorien, die sich bisher zwar schon empirisch „bewährt“ haben, einem härteren Test zu unterwerfen, um sie ggf. weiterzuentwickeln oder zu präzisieren.

Die *Logik* des Hypothesentests ist anhand des *H-O-Schemas* der Erklärung (s.o., Abschnitt 2.4.1) leicht nachzuvollziehen. Vorab gegeben ist die Aussage (1) – die zu prüfende Hypothese. Gesucht sind Situationen, in denen die in der Wenn-Komponente genannten Aspekte als Randbedingungen (2) empirisch vorliegen. In diesen Situationen wird „beobachtet“,¹⁶ ob auch die aus der Dann-Komponente ableitbaren Sachverhalte (3) empirisch auftreten. Diese „Beobachtung“ kann in „natürlichen Situationen“ geschehen (also in der sozialen Realität, wie sie unabhängig von der Forschung existiert) oder in „künstlich geschaffenen Situationen“, die

15 Dies geschieht allerdings häufiger als gerechtfertigt. Nicht selten werden Interpretationen, die erst bei der Datenanalyse ad hoc formuliert wurden, nachträglich zu forschungsleitenden Hypothesen umgedeutet.

16 „Beobachten“ ist hier wieder im weitesten Sinne zu verstehen. (vgl. Abschnitt 1.2.3).

speziell für den Forschungszweck herbeigeführt werden (Experimentalforschung, vgl. Abschnitt 2.4.3).

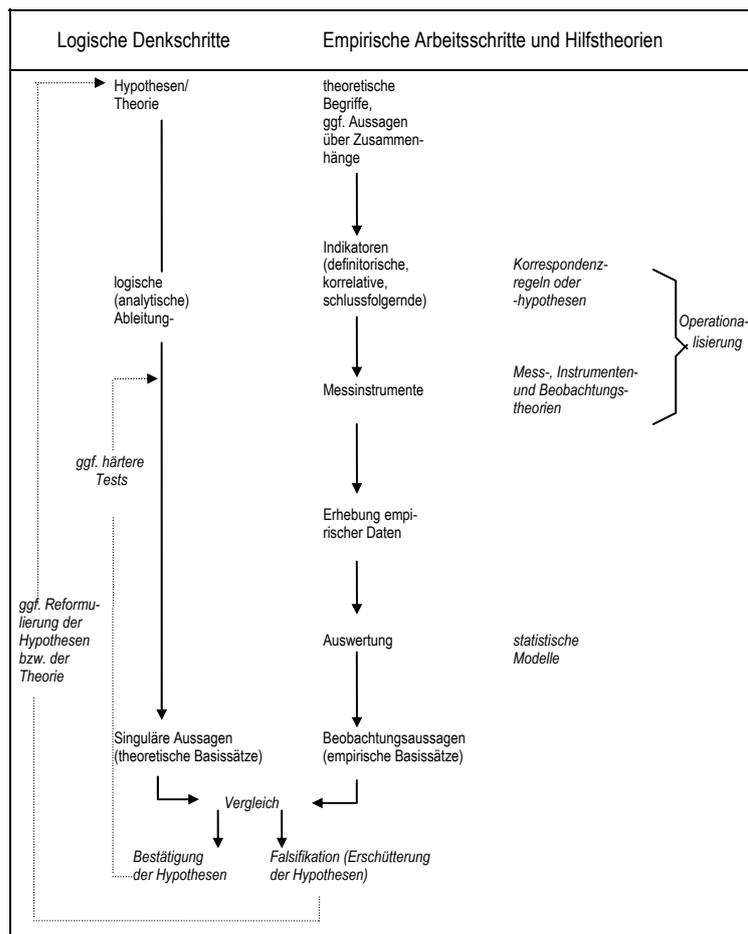
Das *Design* einer hypothesentestenden Untersuchung ist so anzulegen, dass ein gezielter und kontrollierter Vergleich der *empirisch feststellbaren* Sachverhalte mit den aus der Hypothese *ableitbaren* Behauptungen über die empirische Realität möglich ist. Die Überlegungen, die in diesem Zusammenhang anzustellen sind, beziehen sich also noch nicht auf die Wahl bestimmter Instrumente oder Methoden. Mit der Bezeichnung „Design einer hypothesentestenden Untersuchung“ ist vielmehr einerseits eine spezifische Verfahrenslogik (s.o.), andererseits ein spezifischer Argumentationstyp angesprochen.

Der *Argumentationstyp* besteht darin, dass zunächst aus der zu testenden Theorie/Hypothese [(1) im H-O-Schema] deduktiv-logisch Aussagenpaare [(2) und (3) im H-O-Schema] derart abgeleitet werden, dass sie in der Realität beobachtbare Sachverhalte bezeichnen. Anschließend werden diese Deduktionen mit Aussagen über reale (durch empirische Beobachtung festgestellte) Situationskonstellationen verglichen. Stimmen die aus der Theorie/Hypothese abgeleiteten Sätze mit den Beobachtungsaussagen überein, gilt die zu prüfende Theorie/Hypothese als empirisch bestätigt, andernfalls als empirisch widerlegt (zumindest als in ihrem empirischen Geltungsanspruch „erschüttert“).

Wissenschaftstheoretisch sind die aus einer Hypothese ableitbaren Realitätsbehauptungen – also die Gesamtheit der „theorie-implizierten“ Aussagenpaare (2) und (3) – potentielle Konfirmatoren (= „Bestätiger“; falls sie empirisch zutreffen sollten) bzw. potentielle Falsifikatoren der Hypothese (= „Widerleger“; falls sie empirisch nicht zutreffen sollten). Im Folgenden werden sie als hypothesen- bzw. theorie-implizierte Basissätze (oder kürzer: als theoretische Basissätze) bezeichnet.¹⁷

17 Die Bezeichnung „theoretischer Basissatz“ kann man wie folgt verstehen: Die allgemeine Geltung beanspruchende, also *abstrakte* Theorie/Hypothese wird hier *konkret* auf ihre empirische Basis bezogen.

Abbildung: Design-Typ Theorie- und Hypothesentest



Natürlich können in *einer* Untersuchung nicht alle denkbaren (deduktiv ableitbaren) Tatsachenbehauptungen empirisch getestet werden, sondern immer nur eine (gezielte) Auswahl von ihnen: Hypothesen/Theorien sind – so lautet eine Forderung der Wissenschaftstheorie – möglichst

„harten“ Tests zu unterziehen. Hypothesen- bzw. theorietestende Forschung besteht (nach dem Idealbild der kritisch-rationalen Wissenschaftstheorie) also darin, zunächst *solche* Tatsachenbehauptungen zu deduzieren, von denen man annehmen kann, dass sie (möglicherweise) empirisch *nicht* zutreffen.¹⁸ Danach werden in der Realität entweder Situationen geschaffen oder natürliche Situationen aufgesucht, in denen die aus der Wenn-Komponente der Hypothese abgeleiteten Phänomene/Sachverhalte realisiert sind. Zugleich wird mit empirischen Mitteln festgestellt, ob in *diesen* Situationen *auch* die aus der Dann-Komponente abgeleiteten Phänomene / Sachverhalte eintreten bzw. vorliegen. Die *Beschreibungen* der Ergebnisse dieser empirischen Feststellungen sind die *empirischen Basissätze* der Forschung.

Der Vergleich dieser empirischen mit den theoretischen Basissätzen bildet dann die Grundlage der Entscheidung über die Richtigkeit (= Bestätigung) oder Falschheit (= Falsifikation) der zu prüfenden Hypothese bzw. Theorie.

Die Abbildung auf der vorigen Seite fasst das geschilderte Argumentationsmodell theorie- bzw. hypothesentestender Forschung zusammen.

Dieses schlüssig scheinende Verfahren sieht sich allerdings mit zwei gravierenden methodischen Problemen konfrontiert: einerseits dem *Korrespondenzproblem* (das wird in Kapitel 4.3 – Operationalisierung – noch zu vertiefen sein), andererseits dem *Basissatzproblem*.

Das beim Vergleich der theorie-implizierten mit empirischen Basissätzen auftretende *Basissatz-Dilemma* wurde bereits in Abschnitt 1.3.3 skizziert. Es besagt in der jetzt vorgenommenen Präzisierung der Vergleichssituation zwischen „Theorie“ und „Realität“, dass die empirisch gewonnene Beobachtungsaussage „wahr“ sein muss, damit aus ihrem Vergleich mit dem theoretisch abgeleiteten *potentiellen* Falsifikator oder Konfirmator eine Schlussfolgerung auf die „Wahrheit“ der Hypothese begründet werden kann.

Das *Korrespondenzproblem* setzt eine Stufe früher an. Wie schon in Kapitel 1.3 betont, sollen sich die von der analytisch-nomologischen Wissenschaft angestrebten Hypothesen/Theorien durch möglichst hohen Informationsgehalt auszeichnen. Dazu sollen sie nicht nur im Idealfall raum-zeit-unabhängige Geltung beanspruchen, sondern auch in *Begriffen* formuliert sein, die zeitüberdauernd und ortsübergreifend gelten

18 Vielleicht ist es nützlich, wenn Sie jetzt noch einmal das Beispiel in Abschnitt 1.3.6 lesen.

können, also von den Besonderheiten singulärer Situationen bewusst abstrahieren. Außerdem sollen sich die Hypothesen/Theorien nicht auf Aussagen über unmittelbar beobachtbare Phänomene beschränken, sondern auch Vermutungen über nicht direkt „fassbare“ Eigenschaften und Phänomene wie „Vertrauen“ oder „Erreichbarkeit“, über Dispositionen wie „Autoritarismus“ oder „Umzugsbereitschaft“ bzw. über theoretische Konstrukte wie „soziale Schicht“ oder „Rollenkonflikt“ zum Ausdruck bringen.

Mit anderen Worten: Die von der Erfahrungswissenschaft angestrebten Hypothesen sind in „theoretischer Sprache“ formuliert, die Basissätze dagegen (und zwar sowohl die deduktiv abzuleitenden wie die aufgrund von Datenerhebungen zu formulierenden) müssen auf „Beobachtungsbegriffe“ zurückgreifen, d.h. sich auf Sachverhalte beziehen, die unmittelbar wahrnehmbar oder unter Benutzung geeigneter Instrumente erfassbar sind. Zwischen diesen beiden Aussage-Ebenen klafft eine semantische Lücke, die mit Hilfe geeigneter Verknüpfungsregeln zwischen theoretischem Begriff und Beobachtungsbegriffen geschlossen werden muss: Dem theoretisch bezeichneten Phänomen müssen direkt erfahrbare Sachverhalte zugeordnet werden, die als beobachtbare Hinweise (*Indikatoren*) auf das in seiner Allgemeinheit oder Abstraktheit nicht unmittelbar beobachtbare Phänomen dienen können. Diese Verknüpfungsregeln stellen die „Korrespondenz“ (= die Entsprechung) von theoretischer Ebene und Beobachtungsebene her und heißen daher *Korrespondenzregeln*.

Das Problem besteht nun darin, dass nicht immer unbezweifelbare Indikatoren für das theoretisch gemeinte Phänomen benannt und begründet werden können. Während dies bei einem Begriff wie „Erreichbarkeit“ noch auf intuitiver Basis möglich ist, fällt es bei einem Konstrukt wie Rollenkonflikt schon viel schwerer.

Beispiel: Bei einer hypothetischen Aussage über den Zusammenhang zwischen der Erreichbarkeit der Universität von der studentischen Wohnung aus und der Häufigkeit der Anwesenheit der/des Studierenden in der Universität könnten wir z.B. die Entfernung zwischen Wohnung und Hochschule als Indikator nennen. Eine mögliche Korrespondenzregel wäre dann: „Je kürzer der Weg, desto besser die Erreichbarkeit.“ Wir könnten aber auch die Anzahl der Linien des öffentlichen Nahverkehrs und die Dichte der Zug-/Busfolge als Indikatoren nennen. Korrespondenzregeln: „Je mehr Linien des öffentlichen Nahverkehrs die Universität anfahren, desto besser die Erreichbarkeit.“ Und: „Je dichter die Zug- bzw. Busfolge, desto besser die Erreichbarkeit.“

Bei näherem Hinschauen erweist sich die „Korrespondenz“ von Begriff und Indikatoren aber selbst hier als problematisch: Die Zugfolge ist nicht zu jeder Uhrzeit gleich dicht; wir hätten also je nach Tageszeit unterschiedliche „Erreichbarkeiten“. Die Linien des öffentlichen Nahverkehrs nützen dem Studierenden nichts, wenn in der Nähe der eigenen Wohnung keine Haltestelle ist. Die Universität wäre dann zwar „im Prinzip“ gut erreichbar, für den betreffenden Studierenden jedoch mit öffentlichen Nahverkehrsmitteln überhaupt nicht. Wir müssten *für diese Hypothese* die Erreichbarkeit mit Indikatoren operationalisieren, die sich auf die individuelle Situation der/des jeweiligen Studierenden beziehen. Für *eine andere Hypothese* – etwa über einen Zusammenhang zwischen der Erreichbarkeit der Universität und ihrer Bedeutung für die Wirtschaft in der Region – könnten wir allerdings mit *solchen* Indikatoren nichts anfangen, sondern würden ganz andere, nämlich von individuellen Situationen abstrahierende Indikatoren benötigen.

Allgemeiner ausgedrückt: Eine hypothesen- bzw. theorie-*unabhängige* (sozusagen „objektive“) Operationalisierung eines theoretischen Begriffs ist nicht möglich; beides hängt unauflösbar miteinander zusammen.

Selbst wenn das eigentliche Basissatzproblem in einem bestimmten Projekt als vernachlässigbar angesehen werden könnte,¹⁹ wäre aus der Übereinstimmung oder Nicht-Übereinstimmung zwischen theoretischen Tatsachenbehauptungen und empirischen Beobachtungsaussagen kein von der Operationalisierung ablösbarer Schluss auf die Bestätigung oder Falsifizierung der überprüften Hypothese möglich, sondern lediglich auf die Bestätigung oder Falsifizierung der Hypothese *in Verbindung mit den in diesem Projekt verwendeten Korrespondenzregeln*.²⁰ Eine eigentlich falsche Hypothese könnte aufgrund der empirischen Daten bestätigt scheinen (falls die Korrespondenzregeln falsch sind). In gleicher Weise könnte eine eigentlich richtige Hypothese fälschlicherweise empirisch widerlegt erscheinen.

Das Problem ist prinzipiell unlösbar. Der einzige Rat kann nur lauten, die Arbeitsschritte der Operationalisierung (vgl. Kapitel 4) sehr ernst zu nehmen und Korrespondenzregeln nicht lediglich auf Plausibilitätsannahmen, sondern auf empirisch bereits gut bestätigte Hypothesen/Theorien zu stützen – und zwar auf solche Hypothesen/Theorien,

19 Die Messung der gewählten Indikatoren sei mit bewährten und vielfach getesteten Instrumenten geschehen, und die Sorgfalt bei der Datenerhebung und –aufbereitung sei sehr genau kontrolliert worden.

20 Das Korrespondenzproblem (ebenso wie das Basissatzproblem) wird ausführlich in dem bereits genannten Text von H. Esser (1984, I) behandelt.

die nicht in der zu prüfenden Hypothese/Theorie enthalten sind (andernfalls besteht die Gefahr der Tautologisierung der Hypothesenprüfung).

2.4.3 Experiment und Quasi-Experimente

Wird der im vorigen Abschnitt dargestellte „Theorie- bzw. Hypothesentest“ von manchen Autoren als *der* Zweck empirischen Forschens überhaupt eingeschätzt, so gilt das Experiment als *das* Design der Wahl bei jeder Form von Kausalanalysen, d.h. bei *Fragestellungen* des Typs: „Wie und in welcher Stärke wirkt sich X (als Ursache) auf Y (als Folge) aus?“ Im *Lexikon zur Soziologie* wird das Experiment definiert als „planmäßige Beobachtung bestimmter Sachverhalte und ihrer Veränderungen unter vom Forscher kontrollierten und variierten Bedingungen... [Es] unterscheidet sich u.a. dadurch von anderen Beobachtungsformen, dass die beobachteten Vorgänge durch den Forscher hervorgerufen, hergestellt werden.“ (Fuchs-Heinritz u.a., 1994, 190). Besonders in den Naturwissenschaften wird das kontrollierte Experimentieren als eine der zentralen Strategien der Erkenntnisgewinnung, des „Entdeckens“ neuer Zusammenhänge und Gesetzmäßigkeiten in der Natur eingesetzt. In den Sozialwissenschaften stößt die methodisch „reine“ Form des Einsatzes von Experimenten allerdings sehr schnell an ethische Grenzen.

In den Kategorien des *H-O-Erklärungsschemas* (vgl. Abschnitt 2.4.1) lässt sich die experimentelle Fragestellung wie folgt konkretisieren: Der Forscher führt in einer kontrollierten Untersuchungssituation, die insbesondere von allen externen Einflüssen abgeschirmt ist, eine „Maßnahme“ durch – d.h. er realisiert bestimmte „Randbedingungen“: Punkt (2) des *H-O-Schemas* – und setzt seine Untersuchungsobjekte dieser „Maßnahme“ (engl.: *treatment*) aus. Danach beobachtet er, welchen Effekt die „Maßnahme“ auf seine Versuchsobjekte hat, d.h. welche „Wirkungen“ eintreten: Punkt (3) des *H-O-Schemas*. Bei mehrfacher Wiederholung des Experiments unter jeweils gleichen Bedingungen kann aus den Resultaten der Versuchsreihe das in dieser Konstellation zur Geltung kommende Ursache-Wirkungs-Prinzip – Punkt (1) im *H-O-Schema* – abgelesen werden.

Wird in einer Versuchsreihe das „*treatment*“ systematisch variiert, während alle sonstigen Bedingungen konstant bleiben,²¹ kann aus dem

21 Häufig werden Sie für diese Bedingung die Bezeichnung „*ceteris paribus*“ (abgekürzt: c.p.) finden.

Zusammenhang von variierendem „treatment“ und in Abhängigkeit davon variierendem Effekt das Ursache-Wirkung-Prinzip (das empirische „Gesetz“) differenziert ausformuliert werden.

Ging es bei der Konzipierung hypothesentestender Forschung in erster Linie um ein methodologisch abgesichertes *Argumentationsschema*, so konzentrieren sich die methodologischen Überlegungen zum *Design*²² des Experiments um die Gestaltung und *Kontrolle der Untersuchungssituation*.

Im Mittelpunkt des *klassischen Experiments* steht das Bemühen, für die Datenerhebung Bedingungen zu schaffen, in denen *nur* das Ursache-Wirkungs-Prinzip zwischen Maßnahme und Effekt zur Geltung kommen kann. Zum anderen ist Vorsorge zu treffen, dass die Art und die Stärke der vermuteten Kausalwirkungen eindeutig festgestellt und in gültiger Weise – nach Möglichkeit sogar quantitativ – gemessen werden können.

Die *erste Anforderung* kann dadurch erfüllt werden, dass – wie bereits angesprochen – die als (mögliche) Ursache angenommene Einflussgröße (s.o.: „Maßnahme“ oder „treatment“; in der psychologischen Experimentalforschung häufig auch „Stimulus“ genannt) vom Forscher kontrolliert in die Untersuchungssituation eingeführt wird und dass alle sonstigen, im Prinzip denkbaren Einflussgrößen weitestgehend abgeschirmt oder in anderer Weise (z.B. durch „Randomisierung“, s.u.) unwirksam gemacht werden.

Die *zweite Anforderung* ist dann erfüllt, wenn parallel zur Experimentalsituation eine „geeignete“ Vergleichssituation existiert, in der die (angenommene) Ursache *nicht* wirkt, die jedoch in allen anderen Aspekten mit der Experimentalsituation identisch ist.

Im Falle *sozialwissenschaftlicher Experimente* mit Menschen als Untersuchungsobjekten (in der Psychologie meist Versuchspersonen - abgekürzt: Vpn - genannt) zeichnet sich das Design eines „echten“ Experiments dadurch aus, dass es mindestens die folgenden Merkmale aufweist:

- Es existiert eine Experimentalgruppe G1, die dem „treatment“ bzw. dem experimentellen „Stimulus“, also der auf ihre Auswirkungen zu untersuchenden Maßnahme, ausgesetzt wird.

22 Statt *Experimentaldesign* findet man häufig auch die Bezeichnung „Versuchs-anordnung“.

- Es existiert eine in allen wesentlichen Merkmalen äquivalente Kontrollgruppe G2, die dem experimentellen Stimulus nicht ausgesetzt wird, die also vom „treatment“ verschont bleibt.
- In beiden Gruppen werden vor dem Zeitpunkt des „treatments“ und ausreichende Zeit danach die Ausprägungen der abhängigen Variablen (also der Merkmale, bei denen man Auswirkungen durch das „treatment“ erwartet) gemessen.
- Stimmen vor dem „treatment“ in der Experimental- und in der Kontrollgruppe die Verteilungen der abhängigen Variablen überein (was bei äquivalenten Vergleichsgruppen der Fall sein sollte) und sind nach dem „treatment“ Unterschiede zwischen den Gruppen feststellbar, dann werden diese Unterschiede als Effekte interpretiert (d.h. als Auswirkungen der experimentellen Stimuli).

Als Prototyp des sozialwissenschaftlichen Experiments ist 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 Kenntnisse herstellbar ist. Die Zusammensetzung der Gruppen kann in einem solchen Fall *gezielt* vorgenommen werden, die möglichen Einflussgrößen sind gezielt kontrollierbar.

Diese anspruchsvolle Voraussetzung – dass man schon hinreichende empirisch bestätigte Kenntnisse hat, um vollständig kontrollierte Experimente durchzuführen – ist allerdings normalerweise nicht gegeben. Daher machen sich auch so genannte „echte“ sozialwissenschaftliche Experimente den Vorteil des kontrollierten *Zufallsprinzips* zunutze,²³ der darin besteht, auch (noch) unbekannte Merkmale und Faktoren mit angebbarer Wahrscheinlichkeit in einer nicht einseitig verzerrenden Weise zu repräsentieren: Die Versuchspersonen werden „zufällig“ (etwa durch Auslosen) auf Experimental- und Kontrollgruppe verteilt. Dieses Vorgehen wird als „*Randomisierung*“ bezeichnet.

Diese zufällige *Zuordnung* darf *nicht* mit einer *Zufallsauswahl* der Experimentalteilnehmer verwechselt werden – in aller Regel wird man auf Freiwillige angewiesen sein. Das Prinzip der Randomisierung setzt eine Stufe tiefer ein: Sobald eine genügende Anzahl von Versuchspersonen gefunden ist, werden diese zunächst hinsichtlich derjenigen Merk-

23 Auf das Prinzip des „kontrollierten Zufalls“ wird im Detail im Kapitel 6 – Auswahlverfahren – eingegangen.

male, die für den Ausgang des Experiments als bedeutsam gelten – vielleicht Alter, Geschlecht, Bildung –, „geschichtet“ (sozusagen vorgruppiert). *Danach* entscheidet ein Zufallsverfahren, welche Personen aus jeder Schicht der Experimentalgruppe und welche der Kontrollgruppe zugewiesen werden. Auf diese Weise erreicht man zweierlei: zum einen die Bildung unmittelbar äquivalenter Experimental- und Kontrollgruppen hinsichtlich der Schichtungsmerkmale, zum anderen durch das zufällige Zuweisen – das sog. Randomisieren – die Ausschaltung der Gefahr systematischer Ergebnisverzerrungen durch Faktoren, die dem Forscher vorab nicht bekannt sind.

Dieses Zweigruppen-Design (G_1 und G_2) kann noch um zwei weitere Gruppen (eine Experimental- und eine Kontrollgruppe, G_3 und G_4) erweitert werden, in denen man auf *den Vorgang des Messens* vor dem eigentlichen „treatment“ verzichtet. Dadurch wird kontrolliert, ob nicht allein schon *durch die Messung* vor dem „treatment“ Veränderungen in Gang gesetzt wurden (Versuchskaninchen-Effekt).²⁴ Für Untersuchungsgegenstände, bei denen man in verschiedenen Bevölkerungsschichten 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 Bevölkerungsschicht ein komplettes Experimentaldesign).

Die Übersicht auf der folgenden Seite zeigt einige der wichtigsten Experimental-Designs:

24 Gute, kurz gefasste Überblicke über die Experimentalmethode geben *Wellenreuther* 1982, *Frey/Frenz* 1982, *Draper/Pukelsheim* 1996; ausführlich *Zimmermann* 1972; grundlegend zum Experimentalansatz *Holzkamp* 1964 und *Mertens* 1975; aus qualitativer Perspektive *Kleining* 1986.

Übersicht: Design-Strukturen bei Experimenten
(nach Frey/Frenz 1982, 250)

four-group design (Solomon)	<table style="border-collapse: collapse;"> <tr> <td style="padding-right: 5px;">pretest-posttest controlgroup design</td> <td style="padding-right: 5px;">{</td> <td style="padding-right: 5px;">G₁</td> <td style="padding-right: 5px;">R</td> <td style="border-left: 1px solid black; padding-left: 5px;">M₁ × M₂</td> <td rowspan="2" style="padding-left: 10px; vertical-align: middle;">(non)equival. controlgroup design</td> </tr> <tr> <td style="padding-right: 5px;">posttest-only controlgroup design</td> <td style="padding-right: 5px;">{</td> <td style="padding-right: 5px;">G₂</td> <td style="padding-right: 5px;">R</td> <td style="border-left: 1px solid black; padding-left: 5px;">M₁ – M₂</td> </tr> <tr> <td style="padding-right: 5px;"></td> <td style="padding-right: 5px;">{</td> <td style="padding-right: 5px;">G₃</td> <td style="padding-right: 5px;">R</td> <td style="border-left: 1px solid black; padding-left: 5px;">– M</td> <td rowspan="2" style="padding-left: 10px; vertical-align: middle;">static group comparison</td> </tr> <tr> <td style="padding-right: 5px;"></td> <td style="padding-right: 5px;">{</td> <td style="padding-right: 5px;">G₄</td> <td style="padding-right: 5px;">R</td> <td style="border-left: 1px solid black; padding-left: 5px;">× M</td> </tr> </table>	pretest-posttest controlgroup design	{	G ₁	R	M ₁ × M ₂	(non)equival. controlgroup design	posttest-only controlgroup design	{	G ₂	R	M ₁ – M ₂		{	G ₃	R	– M	static group comparison		{	G ₄	R	× M	<table style="border-collapse: collapse;"> <tr> <td style="text-align: center; padding-bottom: 5px;">one-group pretest-posttest design</td> <td style="border-left: 1px solid black; padding-left: 5px;">M₁ × M₂</td> <td rowspan="2" style="padding-left: 10px; vertical-align: middle;">(non)equival. controlgroup design</td> </tr> <tr> <td style="border-left: 1px solid black; padding-left: 5px;">M₁ – M₂</td> <td style="padding-left: 10px; vertical-align: middle;">static group comparison</td> </tr> <tr> <td style="text-align: center; padding-top: 5px;">one-shot case study</td> <td style="border-left: 1px solid black; padding-left: 5px;">× M</td> <td style="padding-left: 10px; vertical-align: middle;">static group comparison</td> </tr> </table>	one-group pretest-posttest design	M ₁ × M ₂	(non)equival. controlgroup design	M ₁ – M ₂	static group comparison	one-shot case study	× M	static group comparison
pretest-posttest controlgroup design	{	G ₁	R	M ₁ × M ₂	(non)equival. controlgroup design																											
posttest-only controlgroup design	{	G ₂	R	M ₁ – M ₂																												
	{	G ₃	R	– M	static group comparison																											
	{	G ₄	R	× M																												
one-group pretest-posttest design	M ₁ × M ₂	(non)equival. controlgroup design																														
M ₁ – M ₂	static group comparison																															
one-shot case study	× M	static group comparison																														
	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border-top: 1px solid black; width: 50%;"></td> <td style="border-top: 1px solid black; width: 50%;"></td> </tr> <tr> <td style="text-align: center;">mit Randomisierung</td> <td style="text-align: center;">ohne Randomisierung</td> </tr> </table>			mit Randomisierung	ohne Randomisierung																											
mit Randomisierung	ohne Randomisierung																															

Erläuterung: G = Gruppe; R = Randomisierung; M = Messung; × = „treatment“.

Geschildert wurde bisher das Design „echter Experimente“. Wie leicht vorstellbar, sind die dafür genannten Bedingungen allenfalls im „Labor“ herstellbar.

Laborexperimente haben jedoch den Nachteil, im Allgemeinen ausgesprochen „realitätsfern“ zu sein. In ihnen wird künstlich eine Situation hergestellt, in der nur ein einziges Merkmal (der experimentelle Stimulus) in kontrollierter Weise wirksam wird und alle anderen möglichen Einflussgrößen „ausgeblendet“ werden (engl.: screening). Dies widerspricht fast allen realen sozialen Situationen, die ja gerade dadurch charakterisiert sind, dass praktisch „alles mit allem“ zusammenhängt. Ideal geeignet sind solche Labor-Untersuchungssituationen für Fragestellungen, die den isolierten Einfluss einer einzigen Variablen für den idealtypischen Fall nachweisen wollen, dass „alle übrigen Bedingungen gleich bleiben“ (sog. *ceteris-paribus*-Klausel), selbst wenn in der Realität dieser Fall nie eintritt. Dies ist etwa für theoretische Argumentationen sowie für den Test und die Konkretisierung von Hypothesen und Theorien von unschätzbarem Vorteil (vgl. das kurze Beispiel in Abschnitt 1.3.6 sowie die ausführliche Darstellung sozialwissenschaftlichen Labor-Experimentaltreihe in *Hunt* 1991, Kapitel 4).

Für andere – insbesondere anwendungsorientierte – Fragestellungen dagegen ist die Realitätsferne ein kaum akzeptabler Nachteil. Anstelle von Laborforschung wird für solche Zwecke *Feldforschung* zu betreiben

sein. Ihr Ziel ist es, soziale Prozesse und Strukturen, wie sie unabhängig vom wissenschaftlichen Forschungsprozess bestehen, zu erfassen und zu analysieren. Versuchen die Forscher dabei, in systematischer Weise die Bedingungen einer gegebenen Feldsituation zu kontrollieren und die für die Untersuchung relevanten Variablen gezielt zu manipulieren, orientieren sie sich am Konzept des *Feldexperiments*. Feldexperimente verfolgen das Ziel, die Logik des klassischen Experiments auch auf Untersuchungsanordnungen im sozialen Feld zu übertragen und dort zu realisieren.

Dies ist in realen sozialen Situationen jedoch fast niemals in vollem Umfang möglich. Beispielsweise wird im allgemeinen nicht für Zwecke der Wissenschaft in die soziale Realität eingegriffen;²⁵ vielmehr muss die Forschung sich an Situationen „anhängen“, in denen im Rahmen des „normalen“ sozialen Handelns Veränderungen in Gang gesetzt werden (etwa Einführung einer neuen Produktionstechnik in der Automobilindustrie, Reform der Verwaltungsabläufe in der Kommune mit dem Ziel „mehr Bürgernähe“, Realisierung eines Arbeitsbeschaffungsprogramms für Langzeitarbeitslose). In solchen Situationen wird aber die Forschung allenfalls geduldet; keinesfalls erhält sie eine federführende Rolle bei der Gestaltung der von ihr untersuchten sozialen Realität.

Insbesondere wäre die Erwartung unrealistisch, Versuchs- und Kontrollgruppen nach von der Forschung vorgegebenen Kriterien zusammenzusetzen und/oder nach dem Zufallsprinzip zu bilden (s.o.: Randomisierung). In manchen Fällen wird es überhaupt nicht möglich sein, über direkte Kontrollgruppen zu verfügen. Häufig sind auch „Vorhermessungen“ der interessierenden Variablen nicht durchführbar. Besondere Probleme verursachen die Kontrolle der „treatments“²⁶ sowie die Abschirmung der übrigen Einflussgrößen.²⁷

Für nicht voll erfüllbare Bedingungen des „echten“ Experiments (s.o.) wird man dann bestrebt sein, Ersatzlösungen zu finden, die (nach

25 Eine Ausnahme bildet der Ansatz der sog. „experimentellen Politik“: Beabsichtigte politische Maßnahmen werden zunächst „versuchsweise“ in begrenztem Rahmen mit wissenschaftlicher Begleitung durchgeführt. Erst aufgrund der so gewonnenen Erfahrungen wird endgültig entschieden (vgl. *Hellstern, G.-M.; Wollmann, H.* (Hg.), 1983: Experimentelle Politik - Reformstrohfeuer oder Lernstrategie?, Opladen).

26 Werden die Maßnahmen – etwa des ABM-Programms – von den zuständigen Stellen auch wirklich so durchgeführt wie vorgesehen und/oder wie nach außen dargestellt?

27 Wie reagiert die nicht direkt betroffene „Umwelt“ auf die Maßnahmen?

Möglichkeit) die gleiche Funktion erfüllen: Anstelle äquivalenter Kontrollgruppen kann man versuchen, Vergleiche mit Situationen anzustellen, in denen die untersuchten Maßnahmen nicht durchgeführt werden (bei dem ABM-Programm etwa Vergleich mit einem anderen Bundesland, in dem die übrigen relevanten Situationsmerkmale ähnlich sind). Anstelle expliziter „Vorher-Messungen“ bei den Versuchspersonen kann man Daten aus früheren Zeiträumen heranziehen. Sind „externe“ Einflüsse auf die Effekte des untersuchten Programms nicht abzuschirmen, müssen diese externen Einflussgrößen gleichfalls gemessen werden, um sie bei der Auswertung der gesammelten Daten berücksichtigen zu können (statistische Variablenkontrolle in der Phase der Datenanalyse anstelle der Situationskontrolle im Experiment).

Untersuchungsanordnungen, die sich an der *Experimentallogik* orientieren, jedoch nicht alle Bedingungen des klassischen Experiments erfüllen können, werden als *Quasi-Experimente* bezeichnet. In der obigen Übersicht sind neben dem „echten“ Experiment einige solcher quasi-experimenteller Designs dargestellt, etwa das Design mit nicht-äquivalenten Kontrollgruppen (fehlende Randomisierung) oder der statische Gruppenvergleich (keine Vorher-Messung).²⁸

2.4.4 Das Evaluationsdesign der Programmforschung

Im alltäglichen Sprachgebrauch bedeutet „Evaluation“ nichts weiter als „Bewertung“. Im Kontext von Forschung wird der *Begriff Evaluation* in zweifachem Sinne benutzt.

Zum einen bezeichnet er eine spezifische *Fragestellung*: Ein bestimmter Sachverhalt oder ein bestimmtes Handeln – z.B. ein politisches Reformprogramm oder eine neue Technologie oder das Handeln einer bestimmten Gruppe, etwa der Hochschullehrer – soll nach vorgegebenen Kriterien mit Hilfe empirischer Informationen bewertet werden.

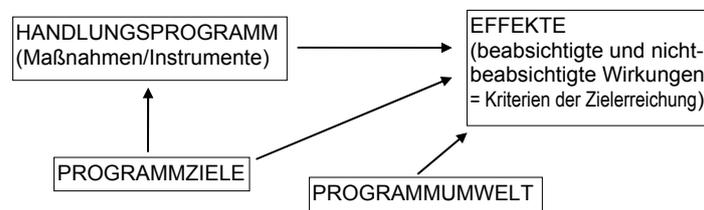
Zum anderen bezeichnet er ein spezifisches *Design*, also einen besonderen Untersuchungsansatz: Das zu evaluierende Handeln – die durchgeführten Maßnahmen sowie die eingesetzten Instrumente, nennen wir es zusammenfassend „Programm“²⁹ – und die durch dieses

28 Für die Darstellung der Methodologie von Quasi-Experimenten sei neben *Frey/Frenz* 1982 auf die „Klassiker“ *Campbell* 1983 sowie *Cook/Campbell* 1979 verwiesen.

29 Unter Programmen versteht die Sozialwissenschaft komplexe Handlungsmodelle, die auf die Erreichung bestimmter Ziele gerichtet sind, auf bestimmten, den

Handeln bewirkten Effekte werden in methodisch kontrollierter Weise miteinander in Beziehung gesetzt und aus der Perspektive der Handlungsziele auf ihren Erfolg hin bewertet. Dabei muss die Forschung explizit berücksichtigen, dass das „Programm“ nur in enger Verflechtung mit seiner sozialen Umwelt durchführbar ist, dass es also von Umgebungseinflüssen nicht abgeschirmt werden kann:

Abbildung: Programmforschung



Beziehen wir Fragestellung und Design wieder auf das in Abschnitt 2.4.1 dargestellte *Hempel-Oppenheim'sche* Erklärungsschema, dann finden wir dessen Komponenten gleich in doppelter Weise wieder.

Zum einen erweist sich das „Programm“ als eine „technologische Aussage“, die ja formal die gleiche Struktur wie eine „Erklärung“ hat. Die Programmziele stellen die angestrebte künftige Situation dar – Punkt (3) im *H-O-Schema*. Die „Maßnahmen“ sind die zu vollziehenden Eingriffe in die gegenwärtigen „Randbedingungen“ – Punkt (2) im *H-O-Schema*. Die Art und Weise, *wie* eingegriffen werden soll, beruht auf Annahmen über Ursache-Wirkungs-Prinzipien – Punkt (1) im *H-O-Schema* –; sie liefern die theoretische Basis des Programms: „*Wenn* die Randbedingungen in bestimmter Weise verändert werden (2), *dann* werden dadurch die beabsichtigten Effekte hervorgerufen (3).“

Zum anderen orientiert sich das *Evaluationsdesign* an eben diesen Komponenten: Es hat sicherzustellen, dass sowohl die existierenden Randbedingungen (2) als auch der Ist-Zustand der Zielvariablen (3) vor Programmbeginn – also zum Zeitpunkt t_0 – empirisch beschrieben werden. Es hat weiter sicherzustellen, dass die *während* der Programmlaufzeit vorgenommenen Eingriffe in die Randbedingungen (2) erfasst wer-

Zielen angemessen erscheinenden Handlungsstrategien beruhen und für deren Abwicklung finanzielle, personelle und sonstige administrative Ressourcen bereitgestellt werden (nach *Hellstern/Wollmann* 1983, 7; grundlegend. *Mayntz* 1980).

den, und zwar sowohl Eingriffe durch die im Programm vorgesehenen *Maßnahmen* als auch andere relevante Veränderungen in der *Programmumwelt* (Diese „begleitende Buchführung“ wird *monitoring* genannt). Das Design hat schließlich sicherzustellen, dass der Zustand der Zielvariablen (3) *nach* Programmdurchführung – also zum Zeitpunkt t_1 – wiederum empirisch beschrieben wird, so dass Art und Ausmaß der *Veränderungen* feststellbar sind.

Ein zentrales methodisches Problem besteht darin, für festgestellte Veränderungen der Zielvariablen (= Unterschiede zwischen den Zeitpunkten t_0 und t_1) zu entscheiden, ob und in welcher Höhe sie durch die Maßnahmen des Programms *bewirkt* wurden oder ob sie auf (programmexterne) Umwelteinflüsse zurückführbar sind. Diese Aufgabe der „*Wirkungszurechnung*“ ist am ehesten lösbar, wenn die Evaluationsforschung sich an der Logik des Feldexperiments orientiert und soweit möglich eine quasi-experimentelle Untersuchungsanordnung mit Vergleichsgruppen realisiert (vgl. Abschnitt 2.4.3).

Der eigentliche *Bewertungsteil* der Evaluationsforschung besteht darin, anhand der festgestellten Effekte den „Erfolg“ des Programms, seine Effektivität einzuschätzen. Als *Maßstab* für die Erfolgswertung gelten die Programmziele, also die *beabsichtigten Effekte*. Die faktisch eingetretenen Wirkungen können in mehr oder weniger hohem Maße den beabsichtigten Effekten entsprechen. Der Programmerfolg ist allerdings nicht schon aus den festgestellten Veränderungen, sondern erst aus der *Wirkungszurechnung* (s.o.) ablesbar. Selbst bei einem absolut wirkungslosen Programm könnten Veränderungen der Zielvariablen in der gewünschten Richtung durch positive Einflüsse der Programmumwelt hervorgerufen worden sein. Andererseits kann der umgekehrte Fall eintreten, dass trotz eines eigentlich wirksamen Programms aufgrund negativer Umgebungseinflüsse nur geringe Veränderungen der Zielvariablen (oder gar Veränderungen in nicht gewünschter Richtung) sichtbar werden. Das Programm selbst war dann trotzdem nicht „erfolglos“, denn es hat immerhin eine Verschlechterung der Situation verhindert. Zusammengefasst: Als „*erfolgreich*“ gilt ein Programm dann, wenn die getroffenen Maßnahmen die Zielvariablen in der gewünschten Richtung und in der gewünschten Stärke beeinflussen.

Evaluation darf sich jedoch nicht auf den Nachweis des Eintretens oder Nicht-Eintretens der beabsichtigten Effekte beschränken. In aller Regel werden neben den vom Programm beabsichtigten auch *nicht beabsichtigte Wirkungen* zu beobachten sein. Diese wiederum können von

positiver oder von negativer Art sein und somit das Gesamturteil über ein Programm in erheblicher Weise beeinflussen. Das Evaluationsdesign muss daher das Untersuchungsfeld von vornherein so weit fassen, dass es über den Eingriffsbereich der Maßnahmen des Programms hinausreicht. Das schließt die Entwicklung eines theoretischen *Wirkungsmodells* – Punkt (1) im *H-O-Schema* – ein, welches auch Alternativen zu den Ursache-Wirkungs-Annahmen berücksichtigt, die dem Programm selbst zugrunde liegen (s.o.). Darüber hinaus impliziert dieser Arbeitsschritt eine fundierte „dimensionale Analyse“³⁰ des Eingriffs- und des Wirkungsfeldes.

Solche umfassenden – auch theoretischen – Vorarbeiten sind erforderlich, da von einer guten Evaluation nicht lediglich Aussagen über den Erfolg oder Misserfolg eines Programms erwartet werden. Vielmehr sollen bei ausbleibendem Erfolg aus den Forschungsergebnissen auch Hinweise auf *Möglichkeiten zur Verbesserung* des Programms ableitbar sein. Dazu ist neben der eigentlichen Wirkungsanalyse u.a. auch die Beurteilung der Programmdurchführung sowie der theoretischen Basis erforderlich.

Selbstverständlich ist eine Evaluation in so umfassender Weise, dass sie alle bisher angesprochenen Aspekte vollständig berücksichtigt, in keinem Projekt realisierbar. Jedes faktische Vorhaben wird Schwerpunkte setzen müssen. Forschungsmethodisch führt dies zu unterschiedlichen *Evaluationstypen*, abhängig davon, welche der in obiger Abbildung „Programmforschung“ angesprochenen Dimensionen und/oder deren Zusammenhänge den Kern der Evaluationsfragestellung bildet.³¹

Eine übliche Unterscheidung bezieht sich auf den Gegenstand der Evaluation. Stehen im Vordergrund die Effekte, die von den Maßnahmen eines Programms hervorgerufen werden, haben wir es mit *Wirkungsanalysen* zu tun. Richtet sich der Blick nicht schwerpunktmäßig auf die Effekte, sondern steht die systematische Untersuchung der Planung, Durchsetzung und des Vollzugs eines Programms und seiner Maßnahmen im Vordergrund, spricht man von *Implementationsforschung*. Von Interesse kann auch sein, ob und in welcher Weise die von einem Handlungsprogramm gebotenen Leistungen von der Zielgruppe, für die diese Leistungen erstellt werden, in Anspruch genommen werden bzw. ob das

30 Dieser Arbeitsschritt wird ausführlich in Kapitel 3.1 und 3.2 behandelt.

31 Ausführlicher dazu: *Kromrey* 1987, 1988, 1995, 2000, 2001, 2003; *Stockmann* 2004.

Angebot auf Zustimmung oder Ablehnung stößt. Für diese Evaluationsrichtung steht der Begriff *Akzeptanzforschung*.

Eine weitere gängige Differenzierung wird hinsichtlich des Zeitpunkts vorgenommen, an dem eine Evaluation ansetzt. Hier kann zwischen einer projektbegleitenden und einer abschließenden Evaluation unterschieden werden. Da üblicherweise bei *begleitender Evaluation* zugleich regelmäßige Rückkoppelungen von Ergebnissen in das Projekt vorgesehen sind, hat die Forschung Konsequenzen für dessen Verlauf; sie wirkt „programmformend“. Ein solches Konzept heißt *formative Evaluation*. Eine erst gegen Ende oder gar nach Abschluss eines Projekts durchgeführte – oder erst dann zugänglich gemachte – Evaluation verzichtet explizit auf formative Effekte. Sie gibt vielmehr im Nachhinein ein zusammenfassendes Evaluationsgutachten ab. Man spricht hier von *summativer Evaluation*.

Schließlich ist noch danach zu unterscheiden, woher die *Kriterien der Evaluation* stammen und wer die *Bewertungsinstanz* ist. Im „traditionellen Fall“ – der hier bisher dargestellt wurde – stammen die Beurteilungskriterien aus dem zu evaluierenden Programm selbst. Seine Implementation sowie seine Wirkungen werden im Lichte seiner eigenen Ziele bewertet. Vorgenommen wird die Beurteilung von den Durchführenden der Evaluationsforschung. Dieser Typ von Beurteilung besteht ausdrücklich nicht in der Abgabe subjektiver Werturteile, sondern in der Formulierung „technologischer Einschätzungen“, die intersubjektiv nachprüfbar sein müssen.

In manchen Fällen wird die eigentliche Bewertung jedoch auf *programm- und evaluationsexterne Instanzen* verlagert. Beispielsweise können Fachgutachten eingeholt werden; oder es werden neutrale Experten befragt, die sich thematisch besonders intensiv mit projektrelevanten Themen befasst haben oder die durch berufliche Erfahrungen mit ähnlich gelagerten Aufgaben ausgewiesen sind.

Unter forschungsmethodischen Gesichtspunkten handelt es sich dabei jedoch nicht um „Evaluationsforschung“, sondern um herkömmliche Umfrageforschung: Unter Verwendung empirischer Instrumente werden Experteneinschätzungen erhoben und ausgewertet. Als eine Variante des Verlagerns der Evaluierung auf eine programmexterne Instanz wird verschiedentlich die *Befragung der Adressaten eines Programms* (Nutzer oder Betroffene) favorisiert. Die Begründung fällt scheinbar leicht: Die Nutzer einer Dienstleistung und/oder die Betroffenen einer Maßnahme sind die „eigentlichen“ Experten. Sie haben den Gegenstand der Unter-

suchung aus eigener Erfahrung kennen gelernt und wissen, wie er (bei ihnen) wirkt. Sie sind also – so wird unterstellt – in der Lage, die konkret-möglichen Urteile darüber zu fällen.

Diese Form von Evaluationsbefragung ist jedoch methodisch besonders problematisch, da die erhobenen Urteile weder den Status von Bewertungen im Sinne „technologischer“ Evaluationen noch von Bewertungen neutraler Experten haben. Es handelt sich um individuell parteiische Werturteile von Personen, die in einer besonderen Beziehung – eben als Nutzer, als Betroffene – zum Untersuchungsgegenstand stehen. Hier sollte nicht von „Evaluation“, sondern von *Akzeptanzerhebung* gesprochen werden.³²

2.4.5 Das deskriptive Survey-Modell: *Querschnitterhebung nicht-experimenteller Daten*

In den vorhergehenden Abschnitten wurden drei wichtige Spezialfälle sozialwissenschaftlicher Forschungsdesigns vorgestellt: der Hypothesen- bzw. Theorietest als der Prototyp wissenschaftlicher Fragestellungen und empirisch-methodologischer Argumentation, die Evaluationsforschung als eine typische anwendungsorientierte Fragestellung sowie das Experiment als der „Königsweg“ methodisch kontrollierter Gestaltung der Untersuchungssituation für die Entdeckung und Prüfung von Ursache-Wirkungs-Zusammenhängen.

Im folgenden Abschnitt geht es nun um den „Forschungsalltag“ und zugleich um die allgemeinste und methodisch umfassendste Untersuchungsanordnung: um das Design für die Erhebung (engl.: survey) und

32 Um diesen Typ von Akzeptanzbefragungen handelt es sich im Allgemeinen, wenn im Hochschulbereich von „Evaluation der Lehre“ die Rede ist. Befragt werden Teilnehmerinnen und Teilnehmer von Lehrveranstaltungen, wie sie die Güte der gebotenen Ausbildung und Betreuung einschätzen. Ausführlich dazu: *Kromrey*, Helmut, 1994: Wie erkennt man „gute Lehre“? Was studentische Vorlesungsbefragungen (nicht) aussagen; in: *Empirische Pädagogik*, 8 (2), 153-168; *ders.*, 1995: Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung; in: *Mohler*, P. (Hg.), 1995: *Universität und Lehre*, Münster, New York, 105-128; *ders.*, 1996: Qualitätsverbesserung in Lehre und Studium statt so genannter Lehrrevaluation; in: *Zeitschrift für Pädagogische Psychologie*, 10 (3/4), 153-166.

Analyse empirischer Informationen zur *Beschreibung und Diagnose* eines interessierenden sozialen Sachverhalts *zu einem Zeitpunkt*.³³

Ziele deskriptiver Untersuchungen können sein,

- als Basis für eine zu treffende Entscheidung empirisch gesicherte aktuelle Erkenntnisse über den in Frage stehenden Gegenstand zu erhalten (Fragestellung: Entscheidungsvorbereitung) oder
- bei unvorhergesehenen gesellschaftlichen Veränderungen umfassende Informationen zur Beurteilung und zum Verständnis dieser Entwicklungen zu gewinnen (Fragestellung: Diagnose) oder
- einen noch relativ unbekanntem empirischen Sachverhalt durch eine möglichst breit angelegte Deskription zu erkunden (Fragestellung: Exploration) oder
- beispielhafte Deskriptionen für eine allgemeine Regelmäßigkeit bzw. „Gesetzmäßigkeit“ zur Verfügung zu stellen (Fragestellung: Illustration) oder
- in regelmäßigen Abständen die Informationen über wichtige Teilbereiche der Gesellschaft zu aktualisieren (Fragestellung: gesellschaftliche Dauerbeobachtung) oder
- zu aktuellen politischen und/oder gesellschaftlichen Themen das Meinungsbild in der Bevölkerung zu erheben (Fragestellung: Meinungsforschung) oder
- die Vorlieben und Konsumneigungen potentieller Kunden für kommerziell anzubietende Güter und Dienstleistungen zu durchleuchten (Fragestellung: Marktforschung)
- und manches andere.

Der *Geltungsbereich* empirischer Deskriptionen kann entweder auf die untersuchten Fälle – bei sog. „Fallstudien“ (vgl. Kapitel 9) sogar auf den untersuchten Fall – beschränkt sein: etwa bei Gegenstandsbeschreibungen zur Vorbereitung spezifischer Entscheidungen oder bei Explorationen. Die Absicht kann aber auch sein, die Befunde über die untersuchten Fälle hinaus zu verallgemeinern: etwa bei Markt- und Meinungsforschungen oder bei gesellschaftlicher Dauerbeobachtung. Während im ersten Fall das Untersuchungsfeld eindeutig abgegrenzt ist, wird im zweiten Fall explizit eine „Grenzüberschreitung“ beabsichtigt: Die Be-

³³ Daher die Bezeichnung „Querschnittserhebung“. Eine „Längsschnittuntersuchung“ bezieht sich auf die Analyse eines Sachverhalts im Zeitverlauf (s. Kapitel 2.1).

funde sollen „repräsentativ“ sein für eine über die untersuchten Fälle weit hinausreichende „Grundgesamtheit“.³⁴

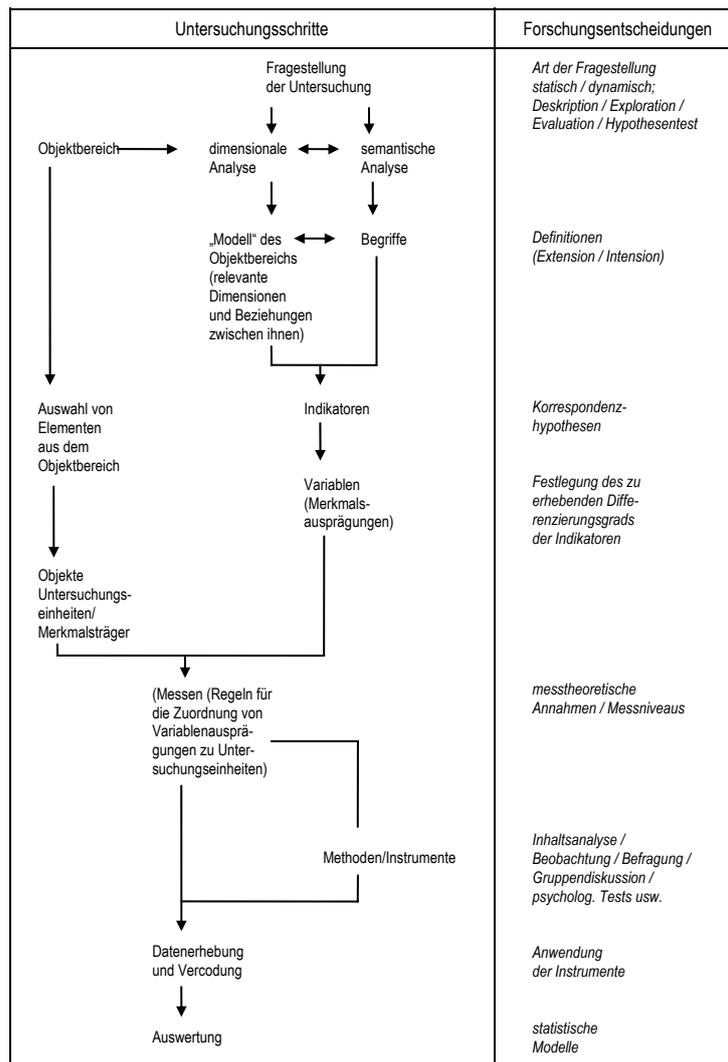
Im Vergleich zu den bisher behandelten Untersuchungsanordnungen verlangt die deskriptive Surveystudie die umfassendsten methodologischen Überlegungen zur Gestaltung des *Forschungsdesigns*. Sie reichen von der Präzisierung der Fragestellung (sowie deren expliziter Einbettung in den Entstehungs- und Verwertungskontext) über die Entwicklung eines untersuchungsleitenden und theoretisch fundierten Gegenstandsmodells sowie die Definition des Untersuchungsfeldes bis hin zur Auswertungsplanung.

Während beispielsweise beim Untersuchungstyp „Theorietest“ die *Fragestellung* durch die zu prüfenden Hypothesen von vornherein eingeschränkt und präzisiert ist, und während dort auch die in Frage kommenden *Objekte der Untersuchung* relativ eindeutig zu bestimmen sind (nämlich mit Hilfe der aus den Hypothesen abzuleitenden „theorieimplizierten Basissätze“, vgl. Kapitel 1.3), erfordern Surveystudien in dieser Hinsicht umfassende Konzipierungsaufgaben mit zentraler Bedeutung für die spätere Brauchbarkeit der Resultate. Während bei Evaluationsforschungen trotz aller Komplexheit der Aufgabenstellung immerhin der *Gegenstand der Untersuchung* (das zu evaluierende Programm und dessen Eingriffsbereich) klar vorgegeben ist, müssen auch in *dieser* Hinsicht bei deskriptiven Studien konzeptionelle Vorleistungen erbracht werden. Diese sollen in ein *Modell des Untersuchungsgegenstands* einmünden, das sowohl dem Gegenstand selbst als auch dem Verwertungszweck der Studie gerecht wird und das zudem durch empirische wie theoretische Vorkenntnisse hinreichend abgesichert ist, um weder bei wichtigen Aspekten Lücken entstehen zu lassen noch Unnötiges und Irrelevantes zu erfassen. Darüber hinaus sollen im Idealfall auch Surveystudien Ansätze für die Weiterentwicklung sozialwissenschaftlicher Theorien bieten, sie sollen nicht „rein deskriptiv“, sondern zugleich „theoretisch relevant“ sein (vgl. Abschnitt 1.3.5).

Das auf der folgenden Seite dargestellte „*Strukturmodell*“ für nicht-experimentelle Forschungsprojekte zeigt die einzelnen Arbeitsschritte und ihre Zusammenhänge auf.

34 Die Methodik repräsentativer Auswahlen wird in Kapitel 6 behandelt.

Abbildung: Strukturmodell nichtexperimenteller Forschung



In Kapitel 2.2 wurde der Forschungsprozess als „eine Reihe ineinander verzahnter Entscheidungen“ in einer Abfolge von elf Etappen dargestellt. Dort wurde schon darauf hingewiesen, dass diese Aneinanderreihung der Arbeitsaufgaben im Zusammenhang eines empirischen Projekts jedoch nicht so (miss)verstanden werden darf, als wäre der Prozess der Forschung eine lineare Folge von Entscheidungsschritten. Auch bei der Betrachtung der Übersicht sollten Sie sich wieder vergegenwärtigen: Es handelt sich um ein *Strukturmodell*, nicht um ein Ablaufmodell.³⁵

Faktisch sind in jedem Projekt alle Teilschritte untereinander vernetzt: Die Projektidee und die erste Formulierung der Forschungsfragestellung – also die „Geburtsphase“ des Vorhabens – ist nicht denkbar, ihre endgültige Präzisierung ist nicht realisierbar ohne ausdrücklichen und intensiven Bezug zur Schlusstappe: zur (gedanklich als Projektziel vorweggenommenen) Aufbereitung und Verwertung der Forschungsergebnisse. Das Projekt wird ja nicht um seiner selbst willen durchgeführt, es soll vielmehr etwas ganz Bestimmtes erreichen. Darauf – auf dieses Projektziel – ist jede Entscheidung zwischen Beginn und Ende des Forschungsprozesses bezogen.

Auch *innerhalb* dieses Prozesses läuft nichts geradlinig ab: Bei der Ausformulierung der untersuchungsleitenden Hypothesen und der Definition der zentralen Begriffe müssen schon die Erhebungsinstrumente mit bedacht werden, müssen zugleich Entscheidungen über die Informanten und Informationsquellen sowie über die „Objekte“ der Forschung getroffen werden (Personen, Familien, Organisationen, soziale Situationen etc.). Bei der Entwicklung und dem Test der Erhebungsinstrumente und Mess-Skalen muss bereits ein grober Auswertungsplan vor Augen stehen (u.a. damit schon bei der Operationalisierung auf das später benötigte Messniveau der Daten geachtet werden kann). Gedanklich muss bei der Konkretisierung des „Designs“ der gesamte Projektablauf vorweggenommen werden. Sobald das Projekt erst einmal „im Feld“ ist, kann bei einem auf Standardisierung angelegten Vorhaben kaum noch etwas „repariert“ werden, was im Zuge der Planung nicht bedacht wurde.

Unter methodischen Gesichtspunkten ist – darauf wurde schon hingewiesen – das hier skizzierte Standardmodell deskriptiver Erhebungen das umfassendste und „kompletteste“ Design der empirischen Sozialfor-

35 Wollte man ein Modell skizzieren, das den Ablauf eines Projekts repräsentiert, wären an vielen Stellen Rückkoppelungsschleifen einzufügen.

schung. Es dient daher in diesem Lehrtext für die folgenden Kapitel als „roter Faden“ der Darstellung und Argumentation.

2.5 Literatur zu Kapitel 2

- Alemann*, Heine von, 1977: Der Forschungsprozeß. Eine Einführung in die Praxis der empirischen Sozialforschung, Stuttgart
- Albert*, Hans, 1972: Wertfreiheit als methodisches Prinzip. Zur Frage der Notwendigkeit einer normativen Sozialwissenschaft; in: *Topitsch*, E. (Hg.): Logik der Sozialwissenschaften, Köln 181-210
- ders.*, 1973: Probleme der Wissenschaftslehre in der Sozialforschung; in: König, R. (Hg.), Handbuch der empirischen Sozialforschung, Bd. 1, Stuttgart (3.Aufl.), 57-103
- ders.*; *Topitsch*, E. (Hg.), 1971: Werturteilsstreit, Darmstadt
- Atteslander*, Peter, u.a., 1991: Methoden der empirischen Sozialforschung (6.Aufl.), Berlin, Kap. 2
- Beck*, Ulrich, 1974: Objektivität und Normativität. Die Theorie-Praxis-Debatte in der deutschen und amerikanischen Soziologie, Reinbek
- Blumer*, Herbert, 1973: Der methodologische Standort des symbolischen Interaktionismus; in: Arbeitsgruppe Bielefelder Soziologen (Hg.): Alltagswissen, Interaktion und gesellschaftliche Wirklichkeit, Bd. 1, Reinbek, 80-146
- Buchhofer*, Bernd, 1979: Projekt und Interview. Eine empirische Untersuchung über den sozialwissenschaftlichen Forschungsprozeß und seine sozio-ökonomischen Bedingungen, Weinheim
- Campbell*, Donald T., 1983: Reforms as Experiments; in: *Struening*, E. L.; *Brewer*, M. B. (eds.): Handbook of Evaluation Research, Beverly Hills, London, 107-137
- Cook*, Th. D.; *Campbell*, D. T., 1979: Quasi-Experimentation: Design & Analysis. Issues for Field Settings, Chicago
- Diekmann*, Andreas, 1995: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen, Reinbek
- Draper*, N. R.; *Pukelsheim*, F., 1996: An Overview of Design of Experiments; in: Statistical Papers, H. 1, 1-32
- Esser*, Hartmut, 1984: Fehler bei der Datenerhebung, Kurs 3604 der Fernuniversität Hagen; Kurseinheit I: Methodologische Probleme der empirischen Kritik von Theorien
- Flick*, Uwe; *Kardoff*, E. von; *Steinke*, I. (Hg.), 2000: Qualitative Forschung. Ein Handbuch, Reinbek
- Frey*, Siegfried; *Frenz*, H.-G., 1982: Experiment und Quasi-Experiment im Feld; in: *Patry*, J.-L. (Hg.): Feldforschung, Bern, 229-258
- Friedrich*, Walter; *Hennig*, W., 1975: Der sozialwissenschaftliche Forschungsprozess. Zur Methodologie, Methodik und Organisation der marxistisch-leninistischen Sozialforschung, Berlin-DDR, Teil B, Kap. 1.1-1.3, 3.3

- Friedrichs*, Jürgen, 1977: Methoden empirischer Sozialforschung, Reinbek, Kap.1, 2.1, 3
- Fuchs-Heinritz*, Werner; *Lautmann*, Rüdiger; *Rammstedt*, Otthein; Wienold, Hanns (Hg.), 1994: Lexikon zur Soziologie, Opladen (3. Aufl.)
- Galtung*, Johan, 1978: Methodologie und Ideologie, Frankfurt/M.
- Hellstern*, Gerd-M.; *Wollmann*, H., 1983: Evaluierungsforschung. Ansätze und Methoden, dargestellt am Beispiel des Städtebaus, Basel, Stuttgart
- Holzkamp*, Klaus, 1964: Theorie und Experiment in der Psychologie, Berlin
- Howard*, K.; *Sharp*, J., 1983: The Management of a Student Research Project, Aldershot
- Hunt*, Morton, 1991: Die Praxis der Sozialforschung: Reportagen aus dem Alltag einer Wissenschaft, Frankfurt/M., New York
- Kleinig*, Gerhard, 1986: Das qualitative Experiment; in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 38 (4), 724-750
- Kriz*, Jürgen, 1981: Methodenkritik empirischer Sozialforschung, Stuttgart
- Kromrey*, Helmut, 1987: Zur Verallgemeinerbarkeit empirischer Befunde bei nichtrepräsentativen Stichproben; in: Rundfunk und Fernsehen, H. 4, 478-499
- ders.*, 1988: Akzeptanz- und Begleitforschung. Methodische Ansätze, Möglichkeiten und Grenzen; in: Massacommunicatie, H. 3, 221-242
- ders.*, 1994: Strategien des Informationsmanagements in der Sozialforschung. Ein Vergleich quantitativer und qualitativer Ansätze; in: Angewandte Sozialforschung, H. 3, 163-184
- ders.*, 1995: Evaluation. Empirische Konzepte zur Bewertung von Handlungsprogrammen und die Schwierigkeiten ihrer Realisierung; in: Zeitschrift für Sozialisationsforschung und Erziehungssoziologie (ZSE), H. 4, 313-336
- ders.*, 2001: Evaluation – ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis; in: Sozialwissenschaften und Berufspraxis, 24 (1), 105-131
- ders.*, 2003: Evaluation in Wissenschaft und Gesellschaft; in: Zeitschrift für Evaluation (ZfEv), H. 1, 93-116
- Lamnek*, Siegfried, 1995: Qualitative Sozialforschung, 2 Bde., München/ Weinheim (3. Aufl.)
- Mayntz*, Renate, 1980: Die Entwicklung des analytischen Paradigmas der Implementationsforschung; in: dies. (Hg.): Implementation politischer Programme, Königstein/Ts., 1-17
- Mertens*, Wolfgang, 1975: Sozialpsychologie des Experiments. Das Experiment als soziale Interaktion, Hamburg
- Popper*, Karl R., 1973: Die offene Gesellschaft und ihre Feinde, 2 Bde., Bern, München
- Riley*, Mathilda, W., 1963: Sociological Research. A Case Approach, New York
- Stappf*, Kurt H., 1984: Laboruntersuchungen; in: Roth, E. (Hg.): Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis, München, Wien

Stockmann, Reinhard (Hg.). 2004: Evaluationsforschung, Grundlagen und ausgewählte Forschungsfelder, Opladen (2. Aufl.)

Wellenreuther, Martin, 1982: Grundkurs: Empirische Forschungsmethoden für Pädagogen, Psychologen, Soziologen, Königstein/Ts., 37-102

Zetterberg, Hans L., 1973: Theorie, Forschung und Praxis in der Soziologie, in: König, R. (Hg.), Handbuch der empirischen Sozialforschung, Bd. 1, Stuttgart (3.Aufl.), 103-160

Zimmermann, Ekkart, 1972: Das Experiment in den Sozialwissenschaften, Stuttgart