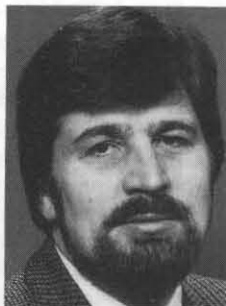


## Methoden und Techniken der Verhaltensuntersuchung 24.

Autoren der Kollegstunde: Walter Bungard / Heiner Ellgring / Peter Vogel  
unter Mitarbeit von Conny Antoni

Autoren des Studienbegleitbriefs: Heiner Ellgring/Peter Vogel/Walter Bungard

Heiner ELLGRING (44), Studium der Psychologie in Marburg und Hamburg. 1975 Promotion, 1984 Habilitation. Ab 1968 wissenschaftlicher Mitarbeiter am Max-Planck-Institut für Psychiatrie, Abteilung Psychologie, in München, mit Unterbrechungen durch Lehrstuhlvertretung an der Justus-Liebig-Universität Gießen. Ab 1979 Leiter der Arbeitsgruppe Sozialpsychologie. Privatdozent am Fachbereich Psychologie der Justus-Liebig-Universität Gießen. Arbeiten zur nonverbalen Kommunikation und Emotionen in Zusammenhang mit psychopathologischen Störungen, Einsatz video-analytischer Verfahren zur Untersuchung und Veränderung des Verhaltens.



Peter VOGEL (39), Studium der Biologie an der Universität Basel. 1979 Promotion. Anschließend wissenschaftlicher Mitarbeiter der Arbeitsgruppe für Verhaltensforschung an der Universität Bochum und Stipendiat der Abteilung WICKLER am Max-Planck-Institut für Verhaltensphysiologie, Seewiesen. Seit 1987 Dozent an der zoologischen Abteilung der University of the West Indies, Jamaika. Arbeitsgebiet: Öko-Ethologie von Räuber-Beute-Systemen.



Walter BUNGARD (41), Studium der Soziologie, Sozialpsychologie, Volkswirtschaftslehre, 1975 Promotion, 1981 Habilitation an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln. Seit 1983 Professor für Wirtschafts- und Organisationspsychologie in Mannheim. Wissenschaftliche Arbeiten aus dem Bereich der Arbeits- und Organisationspsychologie und zur Methodologie der Wirtschafts- und Sozialwissenschaften.



### ***Vor der 24. Kollegstunde zu bearbeiten***

#### **Allgemeine Einführung**

Methoden und Techniken sind in der Regel ein ungeliebter Teil im Studium von Psychologie und Biologie. Warum muß man sich mit Statistik, mit quantitativen Beobachtungsverfahren, mit Experimenten usw. plagen, wenn sich das Verhalten von Mensch und Tier doch so anschaulich schildern läßt? Sofern man kritisch an einen Sachverhalt herangeht, so wird man bald mehrere Gründe entdecken, warum Wissenschaftler ihren Methoden so große Aufmerksamkeit schenken.

Diese Studieneinheit möchte die Struktur methodischen Vorgehens in der psychobiologischen Forschung vorstellen und mit einigen „Werkzeugen“ vertraut machen,

die sich bisher als brauchbar erwiesen haben. Ein wissenschaftliches Vorgehen bei der Untersuchung des Verhaltens zeichnet sich dadurch aus, daß der Ablauf der Untersuchung einer bestimmten Systematik folgt und daß die verwendeten Verfahren verschiedenen Kriterien genügen. Der Ablauf einer Untersuchung orientiert sich dabei vor allem an Hypothesen, die man zunächst entwickelt, dann überprüft und gegebenenfalls revidiert. Das allgemeine Ziel eines solchen Vorgehens ist die Weiterentwicklung von Theorien zur Erklärung des Verhaltens.

Ein fiktives Beispiel wäre die Hypothese, daß Knaben aggressiver auf Fehlschläge reagieren als Mädchen. Die Hypothese entwickelt sich aufgrund von Beobachtungen und von Hinweisen aus früheren Untersuchungen, in denen sich Knaben allgemein als aggressiver zeigten. Auch theoretische Vorstellungen, daß aggressive männliche Individuen eine größere Reproduktionschance haben, könnten zur Hypothese beitragen. Die Hypothese ließe sich überprüfen, indem man Kinder beiderlei Geschlechts gezielt in Fehlschlagsituationen beobachtet oder indem man experimentell Situationen herstellt, in denen die Kinder Fehlschläge erleben müssen. Die Überprüfung könnte ergeben, daß die Knaben zwar im Durchschnitt etwas häufiger ungehalten in den beobachteten Fehlschlagsituationen reagierten, daß dies aber auch bei zahlreichen Mädchen der Fall ist. Aus solch einem Ergebnis könnte man zu dem Schluß gelangen, daß die ursprüngliche Hypothese zu differenzieren ist. Man könnte beispielsweise die Hypothese aufstellen, daß unterschiedliche Erziehungspraktiken für die große Variabilität des beobachteten Verhaltens verantwortlich sind. Nachdem man in dem Beispiel also aufgrund verschiedener Beobachtungen die Hypothese einer geschlechtsspezifischen Aggressionstendenz entwickelt hatte, wurde sie experimentell überprüft und auf Grund der gewonnenen Ergebnisse revidiert. Damit werden gleichzeitig auch die ursprünglich zugrundeliegenden Theorien modifiziert bzw. erweitert.

In jeder der Phasen im Ablauf einer wissenschaftlichen Untersuchung – der Entwicklung von Hypothesen, ihrer Überprüfung, der Bearbeitung gewonnener Ergebnisse und der Gewinnung neuer Hypothesen – kann jeweils eine Vielfalt von Methoden eingesetzt werden. Die Wahl der Methoden richtet sich nach bestimmten Kriterien, vor allem aber nach der Fragestellung und der Art der zu beobachtenden Phänomene.

Bei der *Entwicklung von Hypothesen* kommt es zunächst auf eine sorgfältige Beschreibung des Phänomens an. Bereits hier versucht man, systematisch und strukturiert vorzugehen. Wenn möglich, werden Hypothesen aus bereits bestehenden Theorien abgeleitet. Eine wissenschaftliche Hypothese sollte schließlich so formuliert sein, daß sie auch überprüfbar ist. Die Hypothese, daß männliche Individuen vom Charakter her aggressiver sind als weibliche Individuen, wäre in dieser Form nicht wissenschaftlich überprüfbar. Man müßte zumindest spezifizieren, was man unter „Charakter“ versteht.

Die *Überprüfung von Hypothesen* geschieht in der Psychobiologie zunächst einmal durch Beobachtung. Abhängig von der Art der Fragestellung wählt man die Situationen, in denen man beobachtet (Feldbeobachtung, Experiment), die Population (Art, Grundgesamtheit von Individuen), das Verhalten (z. B. Vogelgesang, Sprache, aggressives Verhalten, nonverbaler Ausdruck usw.) und die Beobachtungsverfahren, mit denen man das Verhalten erfaßt (systematische Beobachtung, Befragung, physiologische und biochemische Messungen usw.). Das menschliche Verhalten erlaubt im Gegensatz zum tierlichen einen Zugang über die Sprache. Ein Untersucher kann eine Person fragen, was sie in einer bestimmten Situation getan hat oder tun würde. Darüber hinaus kann er nach den Gefühlen und Empfindungen fragen, die mit bestimmten Verhaltensweisen verbunden sind. Die Psychologie unterscheidet deshalb zwischen „objektiven“ und „subjektiven“ Verfahren; die Ethologie kennt diese Unterscheidung nicht, da subjektive (sprachliche) Verfahren nicht angewendet werden können. Die Bearbeitung der gewonnenen Verhaltensdaten erfolgt durch statistische Verfahren. Sie dienen dazu, die Beobachtungen und Messungen zusammenzufassen, zu vergleichen und zufallskritisch, das heißt daraufhin zu prüfen, ob die gewonnenen Ergebnisse nicht ein Zufallsprodukt darstellen. Diese zufallskritische Prüfung ist deshalb von besonderer Bedeutung, da man bei

Verhaltensreaktionen immer mit inter- und intra-individueller Variation rechnen muß: Verschiedene Individuen reagieren in einer bestimmten Situation nicht gleich (inter-individuelle Variation), und auch in ähnlichen Situationen reagiert ein Individuum unterschiedlich (intra-individuelle Variation). Um zum Beispiel festzustellen, ob weibliche Personen überzufällig weniger aggressives Verhalten zeigen als männliche Personen, sind die beobachteten Unterschiede durch statistische Verfahren zu prüfen, die solche Variation berücksichtigen. Von Bedeutung für die Theoriebildung sind auch Simulations-Verfahren, mit denen im Computer verschiedene Modelle des Verhaltens theoretisch überprüft werden können.

Am Ende einer Untersuchung stehen die *Interpretation der gefundenen Ergebnisse* und deren theoretische Einordnung, welche zu neuen Beziehungen zwischen Theorien und Hypothesen führen können. Eine Theorie kann durch die Ergebnisse gestützt oder auch geschwächt, manchmal sogar zu Fall gebracht werden. Nahezu immer aber regen die Ergebnisse zu neuen Überlegungen an. Die Erfahrungen, die durch die Untersuchung gewonnen wurden, führen zu neuen Fragen und Hypothesen. Ein Experiment wie unser Beispiel zur Untersuchung einer geschlechtsspezifischen Aggressionstendenz bildet also nicht einen Schlußstein in der Aggressionsforschung, sondern ist eher Ausgangspunkt für weitere Untersuchungen.

Exaktes, quantitatives Erfassen von Verhalten und Situation sowie statistische Auswertung sind heute unabdingbar, sie machen aber nicht selbst die Qualität der Wissenschaft aus. So ist es möglich, mit riesigem apparativen und rechnerischen Aufwand Banalitäten oder Kuriositäten zu untersuchen. Ein statistisch auch noch so gut gesicherter Unterschied, beispielsweise zwischen verschiedenen Gruppen von Individuen, gibt wenig her, wenn man ihn nicht in einen größeren, übergeordneten Zusammenhang stellen kann. Wesentlich gehört es daher zur Methodik einer entwickelten Wissenschaft, Einzeluntersuchungen im Rahmen eines theoretischen Konzeptes zu planen. Entwicklung der Theorie und nicht Anhäufung von Fakten ist das Ziel.

Nach dem Durcharbeiten dieser Studieneinheit sollen Sie in der Lage sein,

Lernziele

- die Bedeutung der Methodik in der Psychobiologie darzustellen;
- verschiedene Untersuchungsstrategien anhand von Beispielen zu erläutern und deren Vor- und Nachteile zu diskutieren;
- die Beschreibung, Quantifizierung, Messung und Interpretation in der Verhaltensanalyse zu erläutern;
- Kriterien für die Qualität von Beobachtungen anzugeben;
- die Möglichkeiten der Kontrolle von unbeabsichtigten Einflüssen (wissenschaftlichen Artefakten) abzuschätzen;
- ethische Probleme der Forschung darzulegen.

## **Gliederung der Kollegstunde**

### *1. Spielregeln methodischen Vorgehens*

Die typischen methodischen Probleme der Psychobiologie werden in dieser Kollegstunde am Beispiel der Stressforschung aufgezeigt: Anhand des bereits in Studieneinheit 15 dargestellten Tierexperiments mit Tupajas werden zunächst die Spielregeln methodischen Vorgehens illustriert. Dabei geht es vor allem um die Logik bei der Einrichtung von Experimental- und Kontrollbedingungen, um die allgemeinen

Voraussetzungen für Kausalschlüsse und um die Fragen der „internen“ und „externen“ Validität.

### 2. Zur Übertragbarkeit der Ergebnisse aus Tierexperimenten auf den Humanbereich

In diesem Abschnitt geht es um das Problem der Übertragbarkeit der Ergebnisse aus Tierexperimenten auf den Humanbereich. Welche Bedeutung haben hierbei die Denkprozesse bei Menschen? Gelten die methodischen Spielregeln auch uneingeschränkt für Humanexperimente? An einem Beispiel, nämlich dem Arbeitsplatz eines Börsenmaklers, wird demonstriert, wie schwierig insbesondere Alltagsbeobachtungen, also Untersuchungen im natürlichen Umfeld von Personen, sind.

### 3. Grundsätzliche Probleme

Im letzten Teil der Sendung werden drei Grundprobleme der psycho-soziologischen Forschung thematisiert: Fragen der „Reaktivität“, Schwierigkeiten bei der Manipulation einzelner Variablen und schließlich ethische Probleme der Forschung.

## Wichtige in der Kollegstunde verwendete Fachausdrücke

**Ethische Richtlinien:** von der Morallehre und den allgemeinen Menschenrechten abgeleitete Handlungsanweisungen mit dem Ziel, die Würde, Integrität sowie die psychische und physische Unversehrtheit des Menschen zu gewährleisten. Bei Untersuchungen an Tieren sind experimentelle Eingriffe auf das absolut notwendige Maß zu beschränken.

**Experimentale Kontrollbedingung:** Auslassung bzw. Veränderung einer bestimmten Bedingung gegenüber dem anderen, sonst gleichartigen Versuch. Aus dem Vergleich der Ergebnisse von beiden Bedingungen kann auf die Wirkung der experimentellen Bedingung geschlossen werden.

**Externe Validität:** Maß für die Verallgemeinbarkeit und Anwendbarkeit der experimentellen Ergebnisse der Forschungsrealität auf die Anwendungsrealität.

**Frequenzanalyse:** Analyse der Anzahl der pro Zeiteinheit auftretenden Elemente, beispielsweise Schwingungen, Augenbewegungen, Worte usw.

**Humanexperiment:** Anwendung experimenteller Methoden auf die Untersuchung menschlichen Verhaltens unter Berücksichtigung ethisch-moralischer Wertmaßstäbe.

**Hypothese:** überprüfbare Behauptung einer Beziehung zweier oder mehrerer Variablen.

**Interne Validität:** Ein Experiment ist dann intern valide, wenn für das experimentelle Ergebnis keine anderen Erklärungen möglich sind als die in der *Hypothese* (s. d.) vorformulierten.

**Inter-Rater-Reliabilität:** Maßzahl für den Grad an Übereinstimmung zwischen den Beurteilungsausprägungen mehrerer Beurteiler über einen Sachverhalt.

**Kausalaussage:** Aussage über die regelmäßige Bedingtheit eines bestimmten, genau umschriebenen Zustandes durch einen vorausgehenden Faktor, der ebenfalls genau umschrieben ist.

**Kognition:** Sammelbegriff für alle Vorgänge oder Strukturen, die mit dem Gewahrwerden und Erkennen zusammenhängen (Wahrnehmung, Erinnerung, Wiedererkennen, Vorstellung, Begriff, Gedanke, aber auch Vermutung, Erwartung, Plan und Problemlösen).

**Objektivität:** Einstellung, die sich streng sachlich an vorhandenen Daten oder Fakten orientiert. In der Psychologie ist Objektivität ein Maß für den Grad an Übereinstimmung, der bei verschiedenen Beobachtungen in der Beurteilung eines Sachverhalts besteht.

**Operationalisierung:** Transformation von Aussagen in empirisch überprüfbare Meßoperationen und genau beschreibbare Sachverhalte.

**Prinzip der Kovariation:** eine von vier Spielregeln, die beim Aufstellen von Kausalhypothesen einzuhalten ist. Die Veränderung einer bestimmten Variablen (Bedingung, Treatment) muß sich in der Veränderung einer anderen Variablen niederschlagen können.

**Randomisierung:** Zufallsauswahl, Auswahl von Individuen durch Verfahren, die jede subjektive Bevorzugung bzw. Vernachlässigung bestimmter Individuen ausschließen (z. B. Auswahl durch Los, nach Zufallsfolge von Ziffern usw.).

**Rating-Skalen:** Skalen zum Zwecke der Selbst- und Fremdbeurteilung. Die Merkmale erscheinen dabei in verschiedene Klassen aufgeteilt (verbal oder durch Abschnitte gekennzeichnet) oder durch Zahlen symbolisiert.

**Reaktivität:** Veränderung dessen, was gemessen werden soll, durch den Vorgang des Messens selbst – weil zum Beispiel Versuchspersonen wissen, daß sie Objekt einer wissenschaftlichen Untersuchung sind.

**Valide Indikatoren:** Hinweisvariablen, die tatsächlich ein Maß für einen behaupteten Sachverhalt sind.

## Während der 24. Kollegstunde zu bearbeiten

### Gliederung der Kollegstunde

#### 1. Spielregeln methodischen Vorgehens (Erläuterungen anhand des Tierexperiments mit Tupajas)

- 1.1. Einführung in the Thematik
- 1.2. Verhaltensbeobachtungen
- 1.3. Formulierung einer Hypothese
- 1.4. Versuchsanordnung: Experimental- und Kontrollbedingungen
- 1.5. Voraussetzungen für Kausalschlüsse
- 1.6. Frage der internen und externen Validität

#### 2. Zur Übertragbarkeit der Ergebnisse aus Tierexperimenten auf den Humanbereich

- 2.1. Die Bedeutung von Denkprozessen bei Menschen
- 2.2. Methodische Spielregeln bei Humanexperimenten
- 2.3. Zur Problematik von Alltagsbeobachtungen im Rahmen der Stressforschung (Beispiel: Arbeitsplatz Börse)
- 2.4. Zusammenfassung

#### 3. Grundsätzliche Probleme

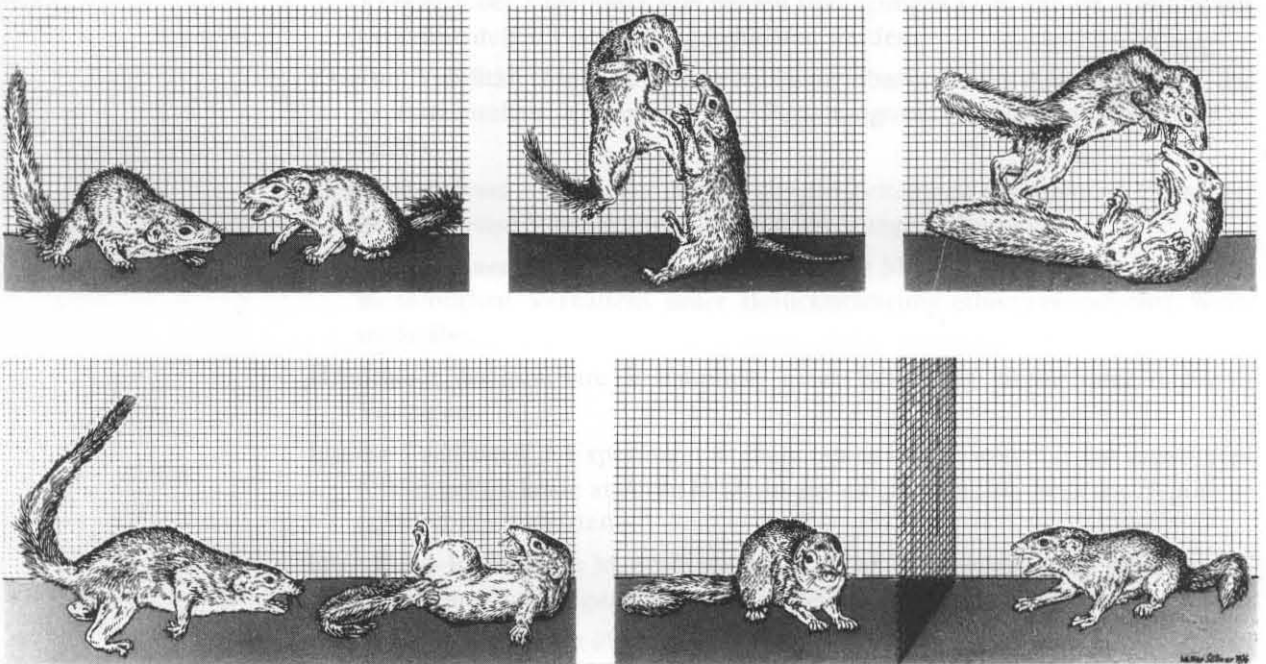
- 3.1. Reaktivitätsprobleme
- 3.2. Schwierigkeiten bei der Manipulation einzelner Variablen
- 3.3. Ethische Probleme der Forschung

### Arbeitsunterlagen

Zu Punkt 1 der Gliederung: „Spielregeln methodischen Vorgehens“

1

Erläuterungen anhand des Tierexperiments mit Tupajas



Zwischen zwei Tupaja-Männchen entbrennt ein Machtkampf um die Herrschaft im Gehege. Der Verlierer des Kampfes, den man durch ein Gitter vom Sieger trennt, stirbt nach wenigen Tagen. Durch systematische und kontrollierte Versuche, die andere Erklärungen nach und nach ausschalten, läßt sich nachweisen, daß die Bedingung dafür der ständige Anblick des überlegenen Tieres und die Angst vor ihm sind.

Zu Punkt 2 der Gliederung: „Zur Übertragbarkeit der Ergebnisse aus Tierexperimenten auf den Humanbereich“

Beispiel: Arbeitsplatz Börse



Verhaltensbeobachtungen an Personen in ihrem natürlichen Umfeld sind schwierig durchzuführen – beispielsweise im hektischen Treiben der Düsseldorfer Börse.

Foto: dpa, Frankfurt.

## Nach der 24. Kollegstunde zu bearbeiten

### Zusammenfassung der Kollegstunde

Im *ersten Teil der Kollegstunde* wurden die Experimente mit den Tupajas, den eichhörnchen-großen Säugetieren aus Südostasien, herangezogen. Die leitende Fragestellung war: Wie kommt man zu einer objektiven Kausalaussage über den Tod des unterlegenen Tupaja-Männchens? Es wurden die vier methodischen Spielregeln hervorgehoben, die, sollen Kausalaussagen gemacht werden, befolgt werden müssen.

- Der Forscher muß die Untersuchungssituation kontrollieren können.
- Der Veränderung einer Variablen muß die Veränderung einer anderen Variablen entsprechen (Prinzip der Kovariation).
- Die Veränderung der vermutlich ursächlichen Variablen muß der Veränderung der vermutlich abhängigen Variablen zeitlich vorausgehen.
- Das Experiment muß prinzipiell beliebig wiederholbar sein.

Im *zweiten Teil der Kollegstunde* wurde zur Beantwortung der Fragen im Zusammenhang von Beobachtungen an Menschen die Reportage über den „gestreßten“ Börsenhändler als Beispiel herangezogen. Es wurde gezeigt, daß man aus der Feldbeobachtung in der Börse sowie aus der Befragung des Börsenhändlers den Streß auf prinzipiell drei verschiedenen Ebenen erfassen konnte: Auf der Verhaltensebene (z. B. Stimme), auf der physiologischen Ebene (z. B. Herzschlag), auf der subjektiven Ebene (Erleben).

Im *dritten Teil der Kollegstunde* wurden drei dieser grundsätzlichen methodischen Schwierigkeiten näher beleuchtet:

- Die „Reaktivitätsproblematik“ (Versuchspersonen wissen, daß sie Objekt einer wissenschaftlichen Studie sind, und stellen je nach Motivation ihr Verhalten ganz unterschiedlich darauf ein).
- Schwierigkeiten bei der Isolierung und Manipulation einzelner Variablen (Menschen nehmen normalerweise eine Situation als Ganze wahr, so daß die Isolierung einer Variablen und deren planmäßige Veränderung in einem Experiment sich notgedrungen als eine qualitative Veränderung der Gesamtsituation auswirken).
- Ethische Probleme der Forschung (Grenzen bei Humanexperimenten auf Grund ethischer Richtlinien).

### 24.0. Methodische Aufgabenstellung der Psychobiologie

#### Grundsätzliches

Ethologie und Psychologie berühren sich auf zwei Ebenen:

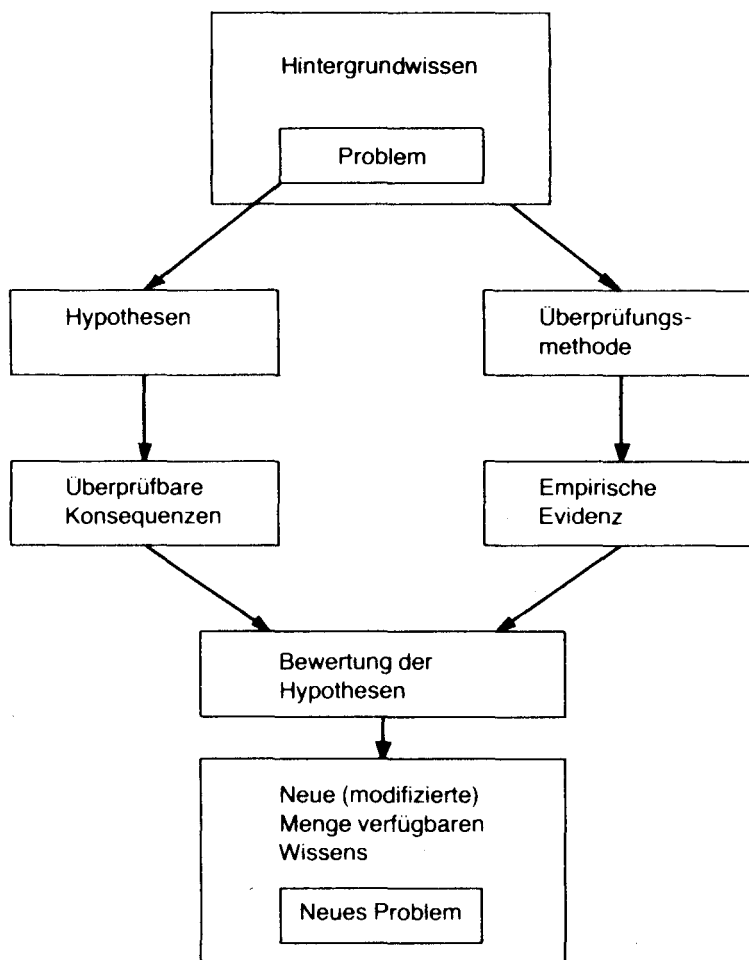
- Beide sind empirische Wissenschaften und beruhen als solche auf einer in Teilen gemeinsamen, allgemeinen Methodologie. Dazu gehört die prinzipielle Struktur des Vorgehens: einer Fragestellung folgen ein Untersuchungsplan, das Messen und Registrieren und die statistische Auswertung und schließlich die Bewertung der Ergebnisse.
- Beide haben einen gemeinsamen Forschungsgegenstand: das Verhalten von Organismen. Sie benutzen daher eine Vielzahl ähnlicher und zum Teil identischer Techniken. Das gilt etwa für die Operationalisierung des Verhaltens (seine Präzisierung als erfaßbare Größe), das Beobachtungsraster und die Meß- und Registrier-einrichtungen.

Auf der verfahrenstechnischen Seite unterscheiden sie sich hauptsächlich darin, daß der Psychologe mit den untersuchten Individuen, den Menschen, sprechen kann. Weiterhin bestehen Unterschiede in den gestellten Fragen oder, anders gesagt, in den Konzepten und Theorien über das Verhalten (vgl. auch STE 6: „Zur Geschichte der Psychobiologie“). Als biologische Wissenschaft stellt die Ethologie die stammesgeschichtliche Natur des Verhaltens in den Vordergrund und operiert mit dem darwinistischen Funktionsbegriff. Ein solcher Ansatz ist der Psychologie weitgehend fremd. Es ist ihr nicht wichtig, ob ein Verhalten die Darwinsche Fitness beeinflußt oder nicht. Ihr Ausgangspunkt sind das Verhalten und das Erleben von Individuen, wobei zur Erklärung der Phänomene Lernprinzipien, Motive, physiologische Zustände usw. herangezogen werden.



In den folgenden Abschnitten sollen wesentliche Teile des methodischen Vorgehens, so, wie sie im Ablauf der Forschung auftreten, behandelt werden (Abb. 1). Dabei nimmt naturgemäß der Bereich der Prüfung von Hypothesen den breitesten Raum ein; denn gerade hierzu wurde das Methodenarsenal entwickelt, das den Verhaltenswissenschaften zur Verfügung steht. Nicht übersehen werden sollten die vielfältigen Probleme, die sich aus der Anwendung solcher Methoden ergeben. Besonders wird auf die Gefahr eingegangen, daß man durch die Beobachtung und Messung die Ergebnisse selbst beeinflußt.

Abb. 1: Grundschrirte auf dem Weg zu erfahrungswissenschaftlichem Erkenntnisgewinn



Ausgangspunkt bildet die Entdeckung eines Problems, das heißt die Feststellung, daß es im Rahmen des verfügbaren (Hintergrund-)Wissens nicht möglich ist, bestimmte Sachverhalte befriedigend zu erklären bzw. vorherzusagen. Der erste Schritt besteht dann darin, vorschlagsweise eine Antwort – eine Hypothese – für die aufgeworfene Forschungsfrage zu formulieren, aus der man überprüfbare Konsequenzen ableitet. In einem nächsten Schritt muß unter Rückgriff auf das Hintergrundwissen nach einer angemessenen Überprüfungsmethode gesucht werden, durch deren Anwendung Daten gewonnen und als empirische Evidenz herangezogen werden können. In einem weiteren Schritt werden dann die aus der Hypothese abgeleiteten überprüfbaren Konsequenzen mit den Daten verglichen, und es wird daraufhin bewertet, ob die Hypothese gestützt oder entkräftet wird. Als Resultat der gesamten Abfolge ergibt sich eine neue (modifizierte) Menge verfügbaren Wissens, die ihrerseits in der Regel zu neuen Problemformulierungen Anlaß gibt.

Beachten sollten man auch die ethischen Probleme, die sich aus dem Einsatz von wissenschaftlichen Methoden bei Mensch und Tier ergeben. Die vergleichsweise kurze Diskussion in dieser Studieneinheit kann dabei der Bedeutung dieses Themas nicht gerecht werden, soll aber zu eigenen Überlegungen anregen.

Wissenschaftliche Untersuchungen kann man als einen fortlaufenden Prozeß von Entscheidungen betrachten: Man entscheidet sich dafür, eine bestimmte Frage zu bearbeiten, und sollte gute Gründe dafür haben. Man entscheidet sich ferner, wie das Problem zu behandeln ist, das heißt, was man beobachten möchte, mit welchen Verfahren an welcher Population und in welcher Situation dies geschehen soll. Schließlich entscheidet man, ob und in welcher Richtung die Beobachtungen weitergeführt haben.

## 24.1. Die Fragestellung

### Schritt 1

Jede ethologische und psychologische Studie hat das Ziel, zu einem allgemeinen Verständnis des Verhaltens der Menschen und Tiere beizutragen. Ausgehend von einem größeren Zusammenhang, stellt der Untersucher eine spezifische Frage und diskutiert seine Ergebnisse im übergeordneten Rahmen. Den übergeordneten Rahmen liefern Theorien, die zu sehr breiten Phänomenen wie der Evolution, dem Lernen usw. aufgestellt wurden oder aber auch spezifische Verhaltensaspekte wie die Aggression erklären wollen. Aus diesen Theorien, die nicht immer ausdrücklich (*explizit*) genannt werden, lassen sich Hypothesen ableiten.

Methoden sind hierbei die Werkzeuge, die eine Überprüfung ermöglichen. Sie lassen sich daher nur im Zusammenhang mit den Theorien und den daraus abgeleiteten Fragestellungen entwickeln und bewerten. Im Alltag nehmen wir meist unsere Vermutungen oder Hypothesen über mögliche Ursachen des Verhaltens als tatsächlich gegeben hin. Wir vermuten zum Beispiel, daß ein Kind das andere schlägt, weil es aggressiv ist, also ein bestimmtes Persönlichkeitsmerkmal oder ein bestimmtes Motiv hat. Diese Vermutung stellen wir zunächst gar nicht weiter in Frage und geben uns mit dieser Erklärung zufrieden. Das Prinzip wissenschaftlicher Untersuchung liegt aber gerade darin, solche Vermutungen oder Hypothesen in Frage zu stellen und alternative Erklärungsmöglichkeiten zu erwägen. Das bedeutet, so zu beobachten, daß eine Widerlegung unserer Theorien und Hypothesen möglich ist. Die Hypothesen müssen dazu so formuliert sein, daß sie grundsätzlich widerlegbar (*falsifizierbar*) sind. Nach POPPER haben nur falsifizierbare Theorien wissenschaftlichen Wert.<sup>1</sup>

Insbesondere aus komplexen Theorien lassen sich sehr verschiedene Hypothesen ableiten. Findet man hierbei widersprüchliche Aussagen, so weist das auf eine Schwäche der Theorie hin. Um die logische Stimmigkeit komplexer Hypothesen zu überprüfen, werden daher auch Computersimulationen der vorhergesagten Ergebnisse herangezogen, so daß die Hypothesen bereits vor der eigentlichen empirischen oder experimentellen Prüfung revidiert werden können.

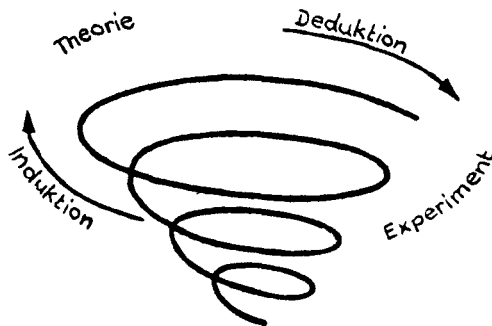
Grundsätzlich können induktive und deduktive Verfahren der Hypothesenbildung (*Abb. 2*) unterschieden werden:

- Bei einem *induktiven Vorgehen* wird gefragt, wie ein bestimmter Verhaltenskomplex aussieht (deskriptive Studie), oder, was unter bestimmten Bedingungen geschieht, ohne spezifische Erwartungen zu nennen. Es können dann Theorien entwickelt werden, die möglichst widerspruchsfrei das Einzelwissen integrieren.
- Bei *deduktivem Vorgehen* werden aus einer Theorie spezifische Voraussagen hergeleitet und gezielt überprüft. Diese Voraussagen sind bis heute vorwiegend

<sup>1</sup> K. R. POPPER: Logik der Forschung. Tübingen 1966.

qualitativer Natur. Die Hypothese lautet, daß gewisse Verhaltensweisen häufiger oder weniger häufig auftreten, wenn sich bestimmte Bedingungen spezifisch verändern.

Abb. 2: Ein Modell für die Beziehungen des Experiments zu den Denkmethoden



Aus: H. SELG: Einführung in die experimentelle Psychologie. Stuttgart <sup>4</sup>1975, S. 12.

Als weiterer Schritt in der Entwicklung der psychobiologischen Wissenschaft wird zunehmend versucht, die Theorien mathematisch zu formulieren. Damit ist es möglich, quantitative Hypothesen zu bilden: Es wird exakt vorausgesagt, wie häufig eine Verhaltensweise unter gegebenen Bedingungen auftreten sollte.

KREBS und DAVIES haben diese Entwicklung am Beispiel des Paarungsverhaltens von Kotfliegen anschaulich dargestellt:

„Vor mehreren hundert Jahren hätte sich ein Naturforscher damit zufrieden gegeben, zwei aufeinander sitzende Kotfliegen zu entdecken, oben das Männchen, unten das Weibchen, und festzustellen, daß die beiden kopulieren. Vor etwa hundert Jahren erkannte Darwin, daß Männchen im allgemeinen um Weibchen konkurrieren. Eine naturgeschichtliche Beschreibung hätte in diesem Stadium auch die Tatsache berücksichtigt, daß die Männchen größer sind als die Weibchen und daß dies eine Folge der sexuellen Selektion sein könnte. Vor rund zehn Jahren wäre dem Evolutionsbiologen als interessant aufgefallen, daß die Männchen nicht nur so lange auf dem Weibchen sitzen bleiben, bis die Spermien übertragen sind, sondern bis zum Ende der Eiablage. Indem das Männchen das Weibchen bewacht, vergewissert es sich, daß seine Spermien nicht von einem nachfolgenden Männchen verdrängt werden. In den letzten Jahren haben die Verhaltensökologen sich an die Frage gemacht, weshalb die Kotfliegenmännchen gerade 40 min. und nicht 10, 20 oder 60 min. auf dem Rücken des Weibchens sitzenbleiben. Bei der Weiterentwicklung einer Theorie, die das ‚Warum?‘ erklärt, wurde deutlich, daß eine entsprechende Analyse auch für nektarsuchende Hummeln, für elterliche Investitionen in die Nachkommen und für viele andere Probleme durchgeführt werden kann. Dieses schrittweise, reduktionistische Fortschreiten von einer breitangelegten Beschreibung zu einer detaillierten, quantitativen Analyse und zu einfachen Verallgemeinerungen ist eine der Hauptentwicklungen der naturgeschichtlichen Forschung.“<sup>2</sup>

Induktive Studien sind gegenüber deduktiven stark zurückgetreten, haben aber auch heute ihre Bedeutung nicht verloren. Wenn die Entwicklung der Theorie den Blick auf Verhaltensweisen öffnet, die bislang wenig Beachtung gefunden haben, können selbst einfache deskriptive Studien wieder ein großes allgemeines Interesse wecken.

Im Folgenden soll je ein Beispiel für deskriptive Studien, für die Überprüfung qualitativer und für die Überprüfung quantitativer Voraussagen gegeben werden:

- Bei sozial lebenden Tieren kommt es gelegentlich vor, daß Erwachsene einzelne Jungtiere ihrer Gruppe töten. Dieses Phänomen fand lange Zeit wenig Beachtung und wurde als eine

<sup>2</sup> J. R. KREBS / N. B. DAVIES: Einführung in die Verhaltensökologie. Stuttgart 1984, S. 331f.

Art krankhafte Abweichung interpretiert. Neuere Überlegungen führten jedoch zu dem Schluß, daß ein solches Verhalten durchaus das Resultat eines Selektionsprozesses sein kann, also eine biologische Funktion besitzt (vgl. STE 4). Führende ethologische Zeitschriften publizieren deshalb heute auch einfache *Beschreibungen* dieser Art.

- GÖTMARK und ANDERSSON beginnen eine Studie mit der Frage, warum manche Vogelarten in Kolonien brüten. Sie stellen dann kurz die zwei wichtigsten Theorien vor: (1) Das Brüten in Kolonien verringert den Raubdruck auf Eier und Jungtiere. (2) Es bewirkt eine erhöhte Effizienz der Nahrungsnutzung; die Kolonien könnten als Informationszentrum dienen. Die Autoren führen dann aus, daß bisherige Untersuchungen, darunter eine eigene, der Informationszentrum-Hypothese keine klare Stütze gaben. Um die Raubdruck-Theorie zu überprüfen, seien Vogelarten am besten geeignet, die teilweise einzeln und teilweise in Kolonien brüten. Eine solche Art ist die Sturmmöwe. Die Autoren stellen schließlich ihre spezifische Frage zur *qualitativen Voraussage* der Hypothese: Überleben Eiattrappen in der Nähe von Kolonien besser als in der Nähe von einzeln brütenden Möwenpaaren? Die Ergebnisse bejahten die Frage.<sup>3</sup>

- Die Theorie der optimalen Nahrungsnutzung sagt voraus, ein Tier sollte sich so verhalten, daß es seine Nahrungsaufnahme pro Zeiteinheit maximiert. (Die Voraussage gilt allerdings nur unter einer Reihe einschränkender Bedingungen, auf die hier nicht näher eingegangen wird.) MILINSKI überprüfte die These in einem spezifischen System: Wird einer Gruppe von Tieren an zwei verschiedenen Stellen Nahrung angeboten; so sollte sie sich so aufteilen, daß alle Tiere gleich viel Nahrung erhalten. Dies ist leicht einsehbar. Ist der Ertrag pro Tier an einer Stelle größer, so sollten Tiere von der anderen Stelle zu dieser überwechseln. Damit sinkt aber der Ertrag pro Tier an der Stelle mit erhöhter Tierzahl und steigt entsprechend an der anderen. Die Wechsel sollten sich fortsetzen, bis der Ertrag pro Tier an beiden Stellen gleich ist. Allerdings müssen verschiedene Voraussetzungen erfüllt sein. So darf es vor allem keinem Tier möglich sein, die anderen von einer Stelle auszuschließen. MILINSKI wählte Stichlinge als geeignete Versuchstiere und Wasserflöhe (Daphnien) einer bestimmten Größe als Futter. Seine spezifische, *quantitative Voraussage* war, daß sich die Fische entsprechend dem Verhältnis der Futter-Eingaberaten zwischen den beiden Stellen aufteilen sollten. Genau dies war der Fall.<sup>4</sup>

In der Formulierung von Fragestellung und Hypothesen wird man sich möglichst auf einen theoretischen Rahmen beziehen und dabei auch die entsprechenden theoretischen Überlegungen, Ableitungen und Prüfpunkte nennen. Dies schließt nicht aus, daß gelegentlich Einzelbeobachtungen zu Fragen anregen, für die eine entscheidbare Hypothese noch nicht formuliert werden kann.

**Aufgabe 1** Erläutern Sie den Unterschied zwischen induktivem und deduktivem Vorgehen in der psychobiologischen Forschung.

.....

.....

.....

.....

.....

.....

.....

.....

.....

## 24.2. Die Untersuchung

**Schritt 2** Eine Reihe von Entscheidungen muß getroffen werden, will man in systematischer Weise die gestellten Fragen empirisch/experimentell untersuchen bzw. seine Hypothesen gezielt überprüfen:

<sup>3</sup> F. GÖTMARK / M. ANDERSSON: Colonial breeding reduces nest predation in the common gull (*Larus canus*). *Animal Behaviour* 32 (1984), S. 485–492.

- *Die Wahl der Stichprobe:* An welchen Personen soll untersucht werden; welche Tierart erscheint geeignet oder notwendig? Ist ein Vergleich mehrerer Arten sinnvoll?
- *Das Erfassen und Setzen von Bedingungen:* Will man das Verhalten in der natürlichen Umgebung beobachten oder experimentelle Bedingungen setzen? Ist es notwendig, die Entwicklung des Verhaltens im Verlauf zu verfolgen, oder reicht eine einmalige Beobachtung aus?
- *Die Erfassung des Verhaltens:* Welche Verhaltensweisen sind für meine Frage von Bedeutung? Wie lassen sie sich beschreiben und registrieren? Wie kann ich die Genauigkeit der Beobachtungen sicherstellen? Für psychologische Untersuchungen sind hierbei die verschiedenen Möglichkeiten der Befragung von Interesse.
- *Die statistische Auswertung:* Um meine Beobachtungen zu ordnen, zu vergleichen und zu bewerten, bedarf es in der Regel statistischer Verfahren. Welche Verfahren sollen oder müssen eingesetzt werden? Bereits in der Planung einer Untersuchung ist es angezeigt, über die spätere Auswertung der Daten möglichst präzise Vorstellungen zu haben. Besonders bei experimentellem Vorgehen sollte die Wahl der situativen Bedingungen mit den später eingesetzten statistischen Verfahren abgestimmt sein.

Ein wissenschaftliches Vorgehen sollte sich dadurch auszeichnen, daß diese Schritte rational getan werden. Auch wenn sich im Laufe einer Untersuchung abweichend vom ursprünglichen Plan Veränderungen ergeben, so spricht das nicht gegen eine solche Systematik. Das wild-fröhliche Drauflosforschen mag zwar attraktiv erscheinen und ist gerade auch in den Phasen der Ideenfindung erfolgreich, es enthebt einen Wissenschaftler jedoch nicht der Notwendigkeit, in einer weiteren Phase die gewonnenen Ideen kritisch und systematisch zu überprüfen.

### Die Wahl der Stichprobe

#### 24.2.1.

Die Auswahl einer geeigneten Population und Stichprobe für die Untersuchung entscheidet mit darüber, wie weit man später seine Ergebnisse und deren Interpretation verallgemeinern (*generalisieren*) kann. So muß man beispielsweise ein genügend breites Spektrum sozialer Schichten berücksichtigen, wenn man Aussagen über die Entwicklung sozialer Verhaltensweisen beim Kinde treffen möchte. Verschiedene Maßnahmen wurden vor allem von den Sozialwissenschaften entwickelt, um repräsentative oder vergleichbare Stichproben zu gewinnen.<sup>5</sup>

Entscheidung 1

Während sich auch für die Psychologie gelegentlich die Auswahl einer geeigneten Tierart als Problem darstellt, ist diese Frage für die Ethologie besonders wichtig. Beide Disziplinen müssen sich aber bei jeder Untersuchung fragen:

- auf welche Gruppe(n) von Individuen (Grundgesamtheit) sie sich beziehen wollen, und
- welche Individuen (Stichprobe) sie daraus für die Beobachtung auswählen wollen.

#### Die Wahl der Bezugsgruppe

##### 24.2.1.1.

Bis heute sind über eine Million Tierarten beschrieben worden. Welche Art sollte ein Ethologe für eine Untersuchung aus dieser Fülle auswählen? Eine erste Möglichkeit wäre, ein Tier zu nehmen, dessen Verhalten noch weitgehend unbekannt ist. Obwohl es genügend solcher Tiere gibt, lohnt sich dieses Vorgehen aber in

Frage 1

4 M. MILINSKI: An evolutionary stable feeding strategy in sticklebacks. *Zeitschrift für Tierpsychologie* 51 (1979), S. 36–40.

5 Vgl. E. K. SCHEUCH: Auswahlverfahren in der Sozialforschung. In: R. KÖNIG (Hrsg.): *Handbuch der empirischen Sozialforschung*. Bd. 3. Stuttgart 1974, S. 1–96.

der Regel nicht. Der Ausgangspunkt ist eine biologische Theorie, und der Untersucher wählt im allgemeinen diejenige Art, die sich zur Überprüfung dieser Theorie am besten eignet. Was eine Art für eine bestimmte Studie geeignet macht, hängt einerseits von der Fragestellung selbst und andererseits von arbeitsökonomischen Faktoren ab.

Als erstes ist zu beachten, daß die gestellte Frage für die gewählte Art überhaupt von Bedeutung ist. Jede Studie geht von bestimmten Annahmen über die Lebensweise einer Art aus, und die untersuchten Tiere müssen diesen Annahmen zumindest annähernd entsprechen. Beispielsweise erforderte MILINSKIS Versuch eine Tierart, bei der Artgenossen sich nicht gegenseitig von einer Freßstelle ausschließen.

Ein weiteres von der Fragestellung bestimmtes Auswahlkriterium betrifft die Bedingungsvariation. Kann ein Untersucher aus irgendwelchen Gründen die ihn interessierenden Bedingungen für ein Verhalten nicht experimentell herstellen, so benötigt er eine Art, in der sie natürlicherweise auftreten. Das Möwenbeispiel illustriert den Fall: die Untersucher wählten eine Art, in der sowohl Kolonie- als auch Einzelbrüter (als Bedingung für die Verteidigungsleistung) vorkommen.

Sind zwei Arten bezüglich der Fragestellung gleich gut geeignet, wird man diejenige wählen, welche die Frage mit geringerem Zeit- und Kostenaufwand beantworten kann. Dieser Aufwand hängt zum einen von der Art selbst ab. Das ideale Tier ist häufig leicht zu beobachten, zu fangen und zu halten, in seinem Aktionsradius beschränkt und vieles andere mehr. Zum anderen hat der Aufwand eine wissenschaftshistorische und bezüglich des Untersuchers eine biographische Komponente. Jede Studie setzt ein hohes Maß an Vorwissen über die untersuchte Art und Artengruppe voraus, sowohl in bezug auf deren Lebensweise als auch auf Arbeitstechniken. Entsprechend haben die Ethologie als Ganzes und ebenso der einzelne Forscher einen Hang zu bestimmten Tiergruppen. Beispielsweise handelten fast ein Drittel der Hauptbeiträge der Zeitschrift *Animal Behaviour* im Jahr 1984 von Vögeln. Der historische Zusammenhang wird deutlich, wenn man bedenkt, daß von den drei Nobelpreisträgern unter den Ethologen – LORENZ, TINBERGEN und VON FRISCH (Abb. 3) – die beiden ersten große Ornithologen sind. Das Übergewicht einzelner Tiergruppen läßt sich nicht vermeiden und ist in allen Zweigen der Biologie vorhanden. Es beschränkt jedoch in manchen Fällen die Möglichkeit, Theorien zu verallgemeinern.

Nicht nur für die Ethologie ist die Wahl der Art von zentraler Bedeutung. Auch in einer psychologischen Untersuchung ist zu entscheiden, welche „Art“ oder besser Gruppe der Spezies *Homo sapiens* geeignet ist. Aus praktischen Gründen ist es vielfach der „Homo studiosus“, meist Student der Psychologie. Mit Recht wurde bei solchen Untersuchungen gefragt, inwieweit sich die an dieser Population gewonnenen Aussagen auf andere Teile der Bevölkerung übertragen lassen. Gerade in der Untersuchung von Emotionen hat man daher in jüngerer Zeit versucht, auch transkulturelle Unterschiede zu berücksichtigen.<sup>6</sup> Das Hauptanliegen der Humanethologie, angeborene Anteile im Verhalten des Menschen zu entdecken,<sup>7</sup> macht es geradezu unabdingbar, daß ein Phänomen an Individuen aus möglichst verschiedenen Kulturen untersucht und verglichen wird.

#### 24.2.1.2. Die Auswahl der zu beobachtenden Individuen

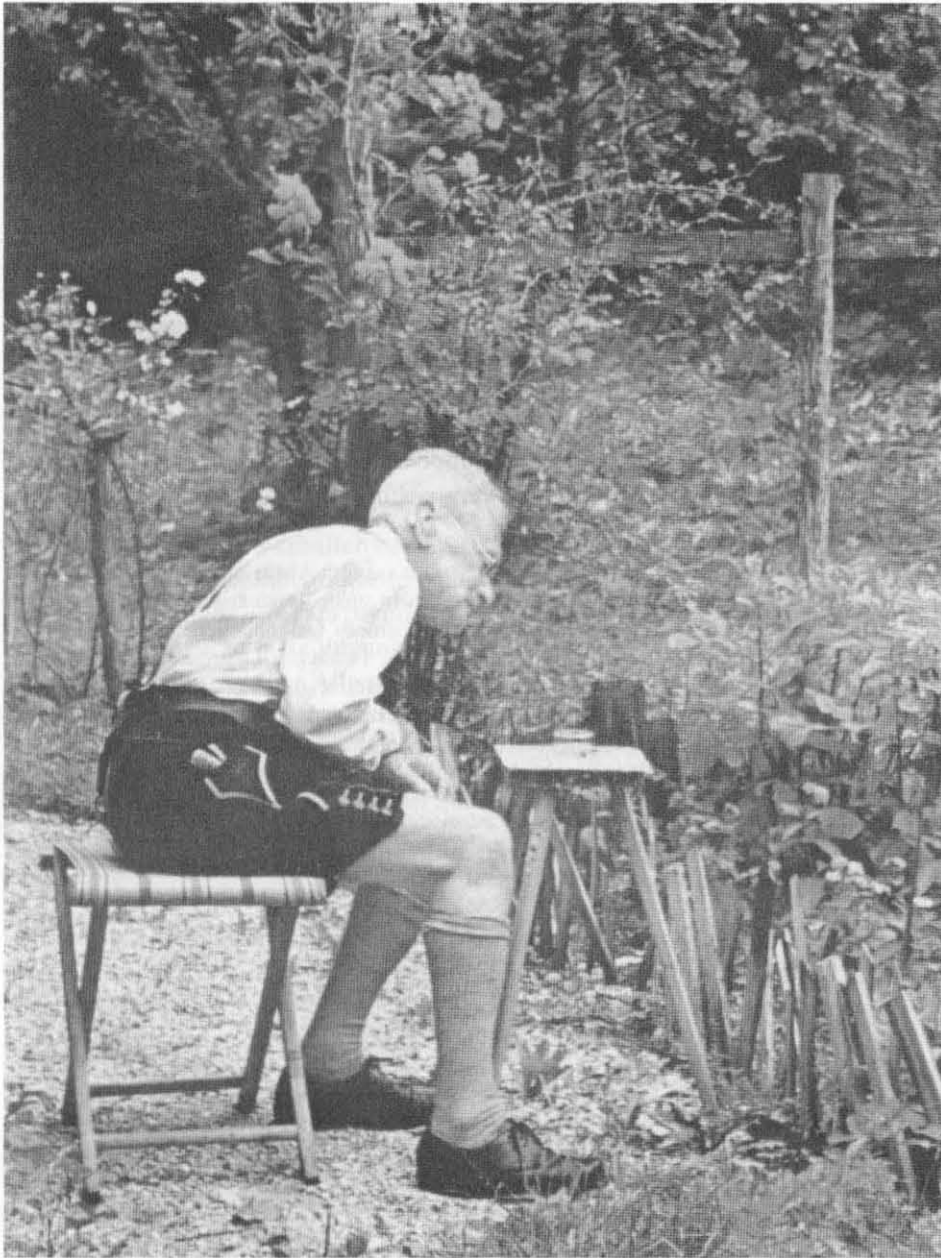
Frage 2

Es ist nicht möglich, sämtliche Individuen einer Art oder Bezugsgruppe in eine Studie einzubeziehen (es sei denn, ein Forscher habe die letzten Exemplare einer

<sup>6</sup> K. R. SCHERER / H. G. WALLBOTT / A. B. SUMMERFIELD (Hrsg.): *Experiencing Emotions. A Cross-cultural Study*. Cambridge 1986.

<sup>7</sup> Vgl. I. EIBL-EIBESFELDT: *Die Biologie des menschlichen Verhaltens. Grundriß der Humanethologie*. München 1984.

Abb. 3: Zur Auswahl der Art



Karl von Frisch wies mit seinen grundlegenden Arbeiten über Sinnesleistungen der Bienen und Fische der Verhaltensforschung neue Wege. Hier testet er das Farbsehen von Bienen.

Aus: N. Tinbergen und die Redaktion der Time-Life Bücher: Tiere und ihr Verhalten. Reinbek 1976, S. 25.

aussterbenden Art vor sich). Ein solches Vorgehen wäre auch völlig sinnlos. Je mehr Individuen bereits untersucht sind, um so geringer wird die Erkenntnis aus der Untersuchung von weiteren. Würden sich sämtliche Individuen in einer bestimmten Situation genau gleich verhalten, könnte man sich mit einem Individuum pro Situation begnügen. Wie viele nötig sind, um ein zuverlässiges Bild zu erhalten, hängt also vom Ausmaß der interindividuellen Variation des untersuchten Verhaltens ab. Grundsätzlich sind um so mehr Individuen (Personen, Tiere) zu prüfen, je größer die Verhaltensunterschiede innerhalb gleichartiger Situationen oder Klassen

von Individuen und je geringer sie zwischen den Situationen oder Klassen sind. Jedoch ist es gerade das Ziel einer Studie, diese Unterschiede herauszufinden, und der Untersucher steht deshalb oft vor der schwierigen Entscheidung, welche Größe seine Stichprobe haben soll. Für den Bereich der Psychologie, der sich mit individuellen Differenzen (von Persönlichkeitsmerkmalen, Reaktionsweisen, Erziehungsstilen usw.) befaßt, werden daher meist größere Stichproben (über 100 Individuen) untersucht. Für andere Bereiche genügen wenige Individuen.

Steht die Anzahl fest, müssen einzelne Individuen für die Stichprobe ausgewählt werden, so daß der Teil das Ganze repräsentiert. Dazu gibt es verschiedene Techniken. Im einfachsten Fall greift man nach einem Zufallsprinzip beliebige Individuen aus der Gesamtpopulation heraus. Will man eine repräsentative Stichprobe untersuchen, so können verschiedene Kriterien bei der Auswahl angelegt werden. Am häufigsten sind es Geschlecht, Alter und sozioökonomischer Status. Besonders beim Vergleich von unterschiedlichen experimentellen Bedingungen ist es notwendig, die verschiedenen Versuchsgruppen nach solchen Kriterien zu „parallelisieren“. So sollen in jeder Gruppe etwa gleich viele männliche bzw. weibliche Versuchspersonen sein, mit vergleichbaren Berufen und in ähnlichem Altersbereich.

Nehmen wir als Beispiel die Untersuchung von MILGRAM zum Autoritätsgehorsam, auf die in Studieneinheit 19 eingegangen wurde.<sup>8</sup> Erinnern wir uns: Versuchspersonen hatten auf Anweisung des Versuchsleiters einen „Schüler“ in einem (angeblichen) Lernexperiment durch elektrische Schläge für seine Fehler zu bestrafen. Tatsächlich wurden allerdings keine Schläge verabreicht, das heißt, der Schüler war ein Helfer des Versuchsleiters und sein Schmerz war simuliert (Abb. 4). Es wurden 4 Gruppen von Versuchspersonen untersucht, die unter verschiedenen Bedingungen zunehmender Nähe zum „Schüler“, dem „Opfer“, die vorgeblichen elektrischen Schläge verabreichen sollten. Es wurde dann festgestellt, wann die Personen den „Gehorsam“ verweigern, das heißt, bis zu welcher Intensität von elektrischen

Abb. 4: Das „Opfer“ in Milgrams Experiment



Aus: K. J. GERGEN / M. M. GERGEN: Social Psychology. New York 1981, S. 360. Copyright 1965 by Stanley Milgram. From the film *Obedience*, distributed by the New York University Film Library.

8 S. MILGRAM: Behavioral study of obedience. *Journal of Abnormal and Social Psychology* 67 (1963), S. 371–378.



Schlägen sie bereit waren, den Schüler zu strafen. In jeder Gruppe nahmen 40 männliche Personen im Alter von 20–50 Jahren teil. Für jede Gruppe galt, daß sie nach Alter und Beruf parallelisiert war: 20% der Personen waren jeweils vom Alter her in den Zwanzigern, jeweils 40% in den Dreißigern bzw. Vierzigern. Jeweils 40% waren Arbeiter bzw. Angestellte, 20% in „Fachberufen“. Hier wurde also versucht, eine *repräsentative Stichprobe* männlicher Personen zu erfassen, die in den verschiedenen Versuchsbedingungen nach Alter und Beruf *parallelisiert* war.

Je repräsentativer eine Stichprobe ist, desto weiter können die Aussagen generalisiert werden. Im genannten Beispiel etwa ist es nicht möglich, das Ergebnis auf die Gesamtbevölkerung zu generalisieren, da nur männliche Personen untersucht wurden. Durch die Parallelisierung wiederum verhindert man, daß ein Ergebnis allein durch die Zusammensetzung der Gruppe beeinflusst wird. Hätte man zum Beispiel in einer Gruppe nur jüngere, in der anderen nur ältere Personen, so könnte der mögliche Unterschied in der Reaktion nicht auf den verschiedenen Versuchsbedingungen, sondern allein auf dem unterschiedlichen Alter beruhen.

### Das Erfassen und Setzen von Bedingungen

### 24.2.2.

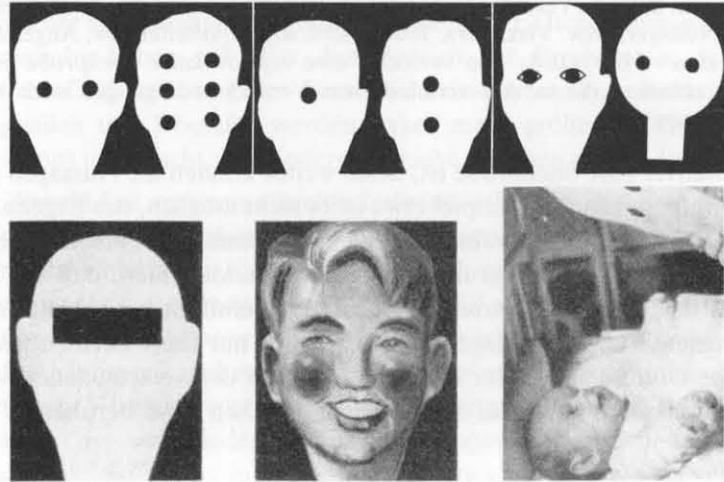
#### Entscheidung 2

Die Bedingungen, die Verhalten beeinflussen können, umfassen die gesamte Innen- und Außenwelt eines Individuums. Wiederum bestimmt die Fragestellung, welche Bedingungen erfaßt oder experimentell gesetzt werden. Eine Aufzählung der möglichen Ebenen von Bedingungen würde sich lesen wie das Inhaltsverzeichnis einer Einführung in die Allgemeine Biologie oder die Psychologie. Je nach Schwerpunkt der untersuchten Bedingungen unterscheidet man verschiedene Zweige der Ethologie, beispielsweise Neuroethologie, Verhaltensphysiologie und Öko-Ethologie. Aus den Teilgebieten der Psychologie sind vor allem die Entwicklungspsychologie und die Sozialpsychologie zu nennen, wobei erstere vor allem die Lern- und Reifungsbedingungen, letztere die sozialen Bedingungen des Verhaltens untersucht.

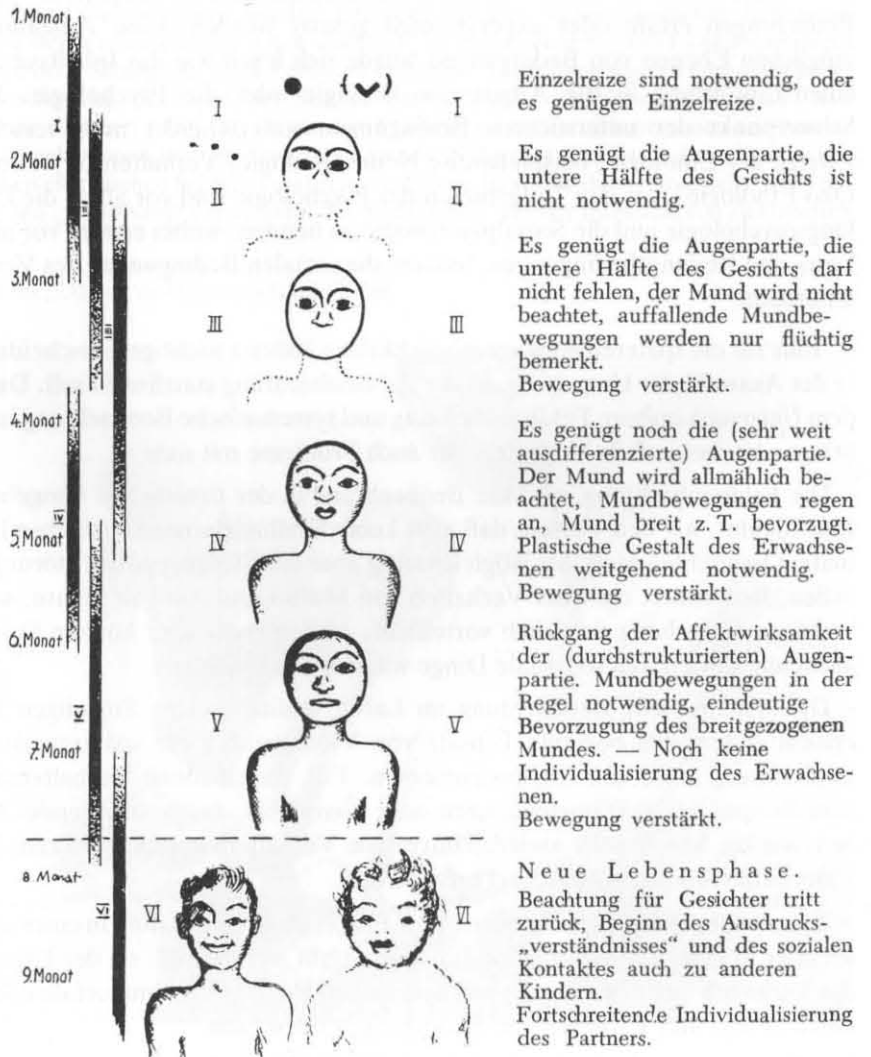
- Eine für die späteren Aussagemöglichkeiten äußerst wichtige Entscheidung liegt in der Auswahl der Umgebung, in der die Beobachtung stattfinden soll. Das heißt, dem *Untersuchungsort*. Feldbeobachtung und systematische Beobachtung im Labor bringen hierbei jeweils Vorteile, aber auch Probleme mit sich:
  - Die Feldbeobachtung, also die Beobachtung in der natürlichen Umgebung des Individuums, hat den Vorteil, daß man keine Einflüsse einer künstlichen Laborsituation befürchten muß. Sie ist gleichzeitig aber anfällig gegenüber Störungen von außen. Beobachtet man das Verhalten von Mutter und Kind zu Hause, so ist die vertraute Umgebung sicherlich vorteilhaft. Andererseits aber können Störeffekte auftreten, etwa durch so banale Dinge wie das Telefonläuten.
  - Die systematische Beobachtung im Labor schließt solche Störungen aus und erlaubt zudem den besseren Einsatz von Videogeräten zur späteren genaueren Beobachtung oder von Meßinstrumenten. Für verschiedene Verhaltensabläufe, zum Beispiel Interaktionssequenzen oder komplexes, rasch ablaufendes Geschehen, wie die Mimik oder andere nonverbale Verhaltensweisen, sind zum Beispiel Videoaufzeichnungen unbedingt erforderlich.
- Ein zweiter Gesichtspunkt betrifft die Frage, ob das Verhalten in einer *ungestörten* oder in einer *experimentellen Situation* erfaßt werden soll: ob der Untersucher das Verhalten unter den gegebenen natürlichen Bedingungen nur erfaßt oder selbst durch Setzen und Variieren von Bedingungen beeinflusst.

Das Experiment erlaubt durch die kontrollierte Veränderung von „unabhängigen Variablen“ eine gezielte Überprüfung von Hypothesen; zudem lassen sich durch „systematische Bedingungsvariation“ Aussagen über kausale Zusammenhänge machen.

Abb. 5: Attrappenversuche von Ahrens



Notwendige, bzw. hinreichende Voraussetzungen zur optischen Auslösung eines Lächelns oder Lachens



Teil der Attrappen und Ergebnisse.

Aus: R. AHRENS: Beitrag zur Entwicklung des Physiognomie- und Mimikerkennens. Teil I und II. *Zeitschrift für Experimentelle und Angewandte Psychologie* 2 (1954), S. 421/445.

Als Beispiel hierfür sollen die Attrappenversuche von AHRENS dienen (Abb. 5), mit denen er die Merkmale bestimmte, die bei einem Säugling das Lächeln auslösen.<sup>9</sup> Während zunächst ein großer schwarzer Punkt als Attrappe ausreicht, ist sehr bald das differenzierte Gesicht der notwendige Stimulus für diese Lächel-Reaktion. Als unabhängige Variable wurde hier die Form der Gesichtsattrappe verändert. Eine weitere unabhängige Variable war das Alter der untersuchten Kinder. Als davon abhängige Variable wurde erfaßt, wie häufig die Lächel-Reaktion jeweils ausgelöst wurde.

- Das *Feldexperiment* kombiniert das Vorgehen der Beobachtung im natürlichen Lebensraum mit einer gezielten Bedingungsvariation, indem in der gewohnten Umgebung der Individuen durch das Handeln des Experimentators ein Ereignisablauf in Gang gesetzt wird, über dessen Abfolge die zu prüfende Hypothese bestimmte Vorhersagen macht.

Ein gewisses Problem besteht darin, wie oft und wie lange die einzelnen Individuen beobachtet werden sollen. Untersuche ich nur zu einem Zeitpunkt, so erfasse ich einen „Querschnitt“ der Population. Gerade für entwicklungspsychologische Studien ist es jedoch notwendig, die Individuen über einen längeren Zeitraum hinweg wiederholt, das heißt im „Längsschnitt“, zu beobachten.

Untersuchungsverfahren und Strategien richten sich also nach der Art des zu untersuchenden Verhaltens und nach den Aussagen, zu denen man durch die Untersuchung gelangen möchte. Für eine generelle Fragestellung – etwa die Entwicklung helfenden Sozialverhaltens bei Kindern – kann es dabei je nach dem Stand der Kenntnis notwendig sein, verschiedene Strategien anzuwenden, das heißt sowohl Feldbeobachtungen anzustellen als auch Experimente durchzuführen.

#### *Der Untersuchungsort*

#### **24.2.2.1.**

Weit vielfältiger als für die Psychologie sind für die Ethologie die Orte, an denen beobachtet und untersucht wird. Dies liegt schon allein an der Vielzahl betrachteter Arten. Allgemein konzentriert sich die Psychologie mehr auf das Labor als auf die natürliche Umgebung.

Wahl 1

Die Untersuchung von Tieren in ihrer natürlichen Umwelt oder im Labor sind nur Pole eines Kontinuums, das alle möglichen Stufen zwischen naturnahen und naturfernen Bedingungen umfaßt. Für gewisse kleine Tiere kann man im Labor Bedingungen schaffen, die dem Freileben sehr ähnlich sind, und in Freigehegen ist es möglich, auch größere Tiere naturnah zu halten. Auf der anderen Seite ist die „freie Wildbahn“ in vielen Fällen nicht mehr mit den natürlichen Bedingungen identisch. Der Mensch hat fast alle Biotope mehr oder weniger stark beeinflußt und damit die Zusammensetzung von Flora und Fauna in weiten Teilen der Welt drastisch verändert. Beispiele dafür sind etwa die Vernichtung großer Teile des tropischen Regenwaldes und die Ausrottung vieler Großraubtiere. Diese Veränderungen können insbesondere für die Untersuchung von funktionellen Fragen ein schwerwiegendes Problem darstellen. Wenn die Umweltfaktoren nicht mehr gegeben sind, die ein bestimmtes Verhalten selektioniert haben, wird es sehr schwierig oder gar unmöglich, dessen adaptiven Wert zu klären.

Die Wahl eines naturnahen oder naturfernen Untersuchungsortes hängt zunächst von der Tierart ab. Je kleiner die Tiere und je geringer ihr Raumanspruch, um so mehr eignen sie sich für das Labor. Im wesentlichen bestimmt jedoch die Fragestellung den Untersuchungsort. Die Feldsituation ist durch eine große Fülle von beeinflussenden Faktoren gekennzeichnet, auf die der Untersucher wenig oder gar keinen Einfluß hat und die er auch nicht alle überwachen kann: ein Räuber taucht auf, es beginnt zu regnen, Spaziergänger wandern quer über die Untersuchungs-

<sup>9</sup> R. AHRENS: Beitrag zur Entwicklung des Physiognomie- und Mimikerkennens. Teil I und II. *Zeitschrift für Experimentelle und Angewandte Psychologie* 2 (1954), S. 412–454.

fläche usw. Viele dieser Faktoren fehlen in der Laborsituation völlig oder sind wesentlich leichter unter Kontrolle zu halten. Außerdem lassen sich die Tiere im Labor meist einfacher beobachten als im Feld. Die Laboruntersuchung sollte unter diesen Gesichtspunkten einer Feldstudie stets vorgezogen werden. Sie findet jedoch in der verhältnismäßig geringen Anzahl der wirkenden Faktoren ihre Grenzen. Für die Erforschung des Verhaltens unter komplexen, interagierenden Bedingungen muß, auf Kosten von präziser Kontrolle und Messung, das Feld gewählt werden. Ein weiterer, wichtiger Punkt kommt hinzu: Das Labor stellt im allgemeinen eine sehr künstliche Umwelt dar und birgt die Gefahr, daß sich Personen und Tiere unnatürlich verhalten. Die gewonnenen Resultate könnten Kunstzeugnisse (*Artefakte*) ohne Bedeutung für die natürliche Situation sein. Ein Untersucher sollte deshalb nach Möglichkeit prüfen, wieweit Laborergebnisse auch für das Freiland gültig (*valide*) sind.

#### 24.2.2.2. *Ungestörte und experimentelle Situation*

Wahl 2

Ist ein Untersucher daran interessiert, wie sich bestimmte Bedingungen auf ein Verhalten auswirken, so kann er entweder ein Individuum oder eine Gruppe in von ihm ungestörter Situation beobachten, bis diese Bedingungen häufig genug eingetreten sind, oder aber sie in einem Experiment gezielt setzen.

In der Untersuchung von MILGRAM wurden vier Bedingungen gesetzt, die den Einfluß der sozialen Nähe zwischen „Lehrer“ und „Schüler“ variierten: „Diese Versuchsreihe umfaßte vier experimentelle Anordnungen. Mit jeder Versuchsbedingung wurde das Opfer ‚psychologisch‘ der Versuchsperson nähergebracht, die ihm die Schocks verabreichte.

Im Rahmen der *ersten* Versuchsanordnung (Fernraum) wurde das Opfer in einen anderen Raum gesetzt und konnte von der Versuchsperson weder gehört noch gesehen werden, außer daß es bei 300 Volt gegen die Wand hämmerte. Über 315 Volt hinaus antwortete es nicht mehr oder wurde nicht länger gehört.

Die *zweite* Versuchsanordnung (Akustische Rückkopplung) war identisch mit der ersten, außer daß stimmliche Proteste eingeführt wurden. Genau wie bei der ersten Versuchsbedingung befand sich das Opfer in einem anliegenden Raum, seine Klagen jedoch konnten klar durch eine leicht geöffnete Tür und durch die Wand des Laboratoriums gehört werden.

Die *dritte* Versuchsanordnung (Raumnähe) war ähnlich der zweiten, außer daß das Opfer nun im selben Raum wie die Versuchsperson plazierte war, in etwa einem halben Meter Entfernung. Auf diese Weise war das Opfer sowohl sicht- wie hörbar; lautliche Zeichen wurden gegeben.

Die *vierte* und letzte Versuchsanordnung dieser Reihe (Berührungsnähe) war gleich der dritten, mit der einen Ausnahme, daß das Opfer einen Schock nur dann erhielt, wenn seine Hand auf einer elektrisierten Platte ruhte. Auf der Stufe von 150 Volt verlangte das Opfer wieder freigelassen zu werden und weigerte sich, bei dieser Versuchsbedingung, die Hand weiter auf die Schockplatte zu legen. Der Versuchsleiter befahl hier der naiven Versuchsperson, des Opfers Hand mit Gewalt auf die Platte zu zwingen. Gehorsam erforderte somit unter dieser Versuchsbedingung, daß die Versuchsperson physischen Kontakt hatte, um ihm eine Strafe über die 150-Volt-Stufe hinaus zu geben.“<sup>10</sup>

Das Ergebnis zeigte, daß die „Lehrer“ um so früher die Bestrafung des „Schülers“ trotz Aufforderung des anwesenden Versuchsleiters abbrachen, je näher sie dem „Opfer“ waren.

Hier wurden also vier Stufen des Merkmals „Soziale Nähe“ hergestellt, um dessen Einfluß auf das Verhalten „Autoritätsgehorsam“ zu überprüfen.

Ein Experiment ist in vielen Fällen der Beobachtung einer ungestörten Situation vorzuziehen und manchmal die einzige Möglichkeit für die Untersuchung. Wichtige Gründe dafür sind:

- Die interessierenden Bedingungen treten in manchen Fällen in der ungestörten Situation nur selten oder zum Teil gar nicht auf.

Viele Forscher haben sich beispielsweise Gedanken darüber gemacht, warum die Zahl der Eier, welche Weibchen einer Vogelart legen, eng um einen Mittelwert schwankt. Eine Theorie lautet, daß die durchschnittliche Eizahl die maximale Zahl an flüggen Jungtieren

<sup>10</sup> Nach S. MILGRAM: Some conditions of obedience and disobedience. *Human Relations* 18 (1965), S. 61.

garantiert. Legt das Weibchen zu wenig Eier, sind die Eltern „unterbeschäftigt“ und könnten bequem mehr Junge hochziehen, legt es dagegen zuviel, stirbt ein Großteil der Jungen, weil die Eltern überfordert sind. Um die Theorie zu prüfen, muß der Aufzucherfolg von Eltern mit unterschiedlicher Gelegegröße ermittelt werden. Da nun aber Eltern mit besonders wenig bzw. viel Eiern selten sind, kann es sich lohnen, die Eizahl experimentell zu verändern: aus gewissen Nestern werden Eier entfernt, in andere hinzugegeben.

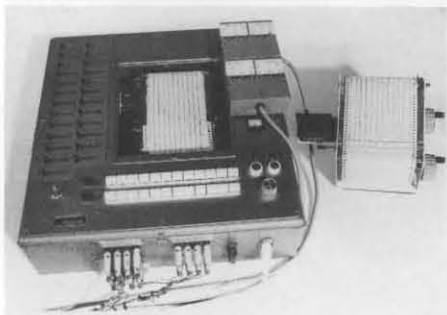
- Es kann unter Umständen technisch zu schwierig, zu kostspielig oder zeitraubend sein, die interessierenden Bedingungen in der ungestörten Situation zu erfassen.

Dies gilt zum Beispiel für viele Untersuchungen zur chemischen Kommunikation. Oft ist es wesentlich einfacher, eine bestimmte chemische Substanz in einem Experiment gezielt freizusetzen, als deren Vorhandensein bzw. Fehlen in der ungestörten Situation nachzuweisen. Auch eine Beobachtung der Interaktion zweier Personen ist wesentlich einfacher im Labor durchzuführen als zu Hause. Geräte wie Videorecorder, Spiegel, Verhaltensschreiber usw. sind im Labor fest installiert und müssen nicht für jede Beobachtung neu aufgebaut werden (Abb. 6).

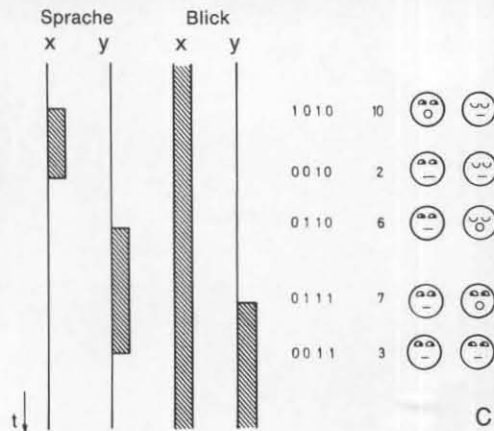
Abb. 6: Laboruntersuchung



A



B



A Integriertes Video-Computer-System zur Analyse nonverbalen Verhaltens. B Mechanischer Verhaltensschreiber (Peiseler) zur simultanen Registrierung verschiedener Merkmale. C Binäre Codierung von Sprechen und Blickzuwendung zweier Personen (von links nach rechts: „An-Aus-Muster“ der vier Verhaltenskanäle, Binär-Code, Dezimal-Code, piktografische Darstellung der Zustände).

- In der ungestörten Situation sind die interessierenden Bedingungen häufig mit anderen eng verknüpft.

Beispielsweise versammeln sich die Männchen verschiedener Froscharten zur Laichzeit an bestimmten Stellen von Gewässern und geben laute Konzerte, während die Weibchen sich diesen Stellen aus oft beträchtlicher Distanz nähern. Ist es der Ruf der Männchen, der sie anlockt, oder bevorzugen sie bestimmte Orte – an denen sich Männchen nur deshalb zusammengefunden haben, weil sie dort am ehesten auf Weibchen treffen? Die Bedingung „Rufen“ ist mit der Bedingung „Ort“ verknüpft, und ein experimenteller Ansatz ist nötig, um die beiden getrennt zu prüfen. Weibchen können im Labor getestet werden, ob sie bevorzugt in Richtung eines Lautsprechers wandern, der den „Gesang“ der Männchen ertönen läßt.

Trotz der vielen Vorteile eines Experimentes sind Beobachtungen in der ungestörten Situation von großer Bedeutung:

Abb. 7: Feldbeobachtung



Jane und Hugo van LAWICK-GOODALL.

- Das Experiment setzt ein breites Basiswissen über das Verhalten von Personen oder die Lebensweise einer Tierart voraus, das nur durch intensives Beobachten in der ungestörten Situation gewonnen werden kann.

Für den Versuch mit den Froschweibchen muß beispielsweise bekannt sein, daß die Männchen der zu untersuchenden Art sich tatsächlich an bestimmten Stellen versammeln und die Weibchen sich diesen Stellen nähern, ebenso Beginn und Dauer der Laichzeit sowie manches mehr. Ohne das Basiswissen besteht die große Gefahr, ein Tier Bedingungen auszusetzen, die an seiner Realität vorbeigehen. Auch das Beobachten von Säuglingen setzt voraus, daß man zum Beispiel das Verhaltensrepertoire kennt oder auch weiß, für wie lange Zeit man Aufmerksamkeit erwarten kann.

- In verschiedenen Fällen ist ein experimenteller Eingriff nur beschränkt möglich. So können die interessierenden Bedingungen zu komplex sein, um sie experimentell



herzustellen, oder ein experimenteller Eingriff könnte das zu untersuchende System als Ganzes empfindlich stören.

Letzteres gilt in besonderem Maße für soziale Gruppen mit individualisierten Beziehungen zwischen den einzelnen Individuen. Jane van LAWICK-GOODALL benötigte Jahre, um das Vertrauen einer Gruppe freilebender Schimpansen zu gewinnen und die basalen sozialen Beziehungen zwischen den Individuen herauszufinden.<sup>11</sup> Sie konnte keine größeren experimentellen Eingriffe riskieren und damit den Zusammenhalt der Gruppe oder deren Vertrauen zu ihr aufs Spiel setzen, lediglich um eine einzelne Hypothese zu prüfen (Abb. 7).

Laboruntersuchungen sind fast immer experimentell ausgerichtet, während in Feldstudien oft die ungestörte Situation im Vordergrund steht. Abhängig von der Fragestellung, von dem zu untersuchenden Phänomen und von dem jeweiligen Kenntnisstand wird man sein Vorgehen wählen – sei es experimentell, sei es beobachtend in der natürlichen Umgebung. Beide Zugangsweisen sind keine sich ausschließenden Alternativen, sondern ergänzen sich.

### 24.2.2.3. *Das Feldexperiment*

Wahl 3

Das Feldexperiment verbindet die Vorteile einer Untersuchung in der natürlichen oder naturnahen Situation mit den Vorteilen einer gezielten Veränderung spezifischer Bedingungen.

Ein solcher methodischer Ansatz ist unabdingbar für viele funktionale Studien, wenn es also um den Nachweis geht, daß ein bestimmtes Verhalten die Fitness eines Tieres beeinflußt. Feldexperimente haben daher in den letzten Jahren für die Ethologie zunehmende Bedeutung erlangt. Drei Beispiele wurden bereits vorgestellt: das Ersetzen von Kohlmeisenmännchen durch Lautsprecher, das Aussetzen von Eiattrappen in einer Möwenkolonie und das Verändern der Gelegegröße.

Für die Untersuchung menschlichen Verhaltens ergeben sich durch das Feldexperiment ebenfalls interessante Zugangsweisen.

So läßt sich beispielsweise das Territorialverhalten des Menschen studieren, indem eine instruierte Person bewußt anderen zu nahe tritt, also soziale Normen durch ihr Verhalten verletzt. Ein anderes Vorgehen besteht darin, einen Platz, etwa in einer Bibliothek, zu „markieren“, indem Dinge wie Hefte oder Taschen gezielt hingelegt werden. Die Frage dabei ist beispielsweise, inwieweit sich andere an solche unsichtbaren Grenzen eines Territoriums halten.<sup>12</sup>

Der Beobachter muß also nicht immer nur eine passive Rolle einnehmen; je nach Untersuchungsgegenstand kann es erforderlich sein, daß er aktiv in den Forschungsprozeß insofern eingreift, als er bestimmte Ausgangsbedingungen schafft, um so (nach wie vor unauffällig) das ihn interessierende Verhalten beobachten zu können. Helfer des Versuchsleiters können beispielsweise bestimmte Reaktionen zeigen, um zu überprüfen, ob andere Personen dieses Verhalten imitieren – eine Versuchsanordnung, die in der Konformitätsforschung häufig Anwendung gefunden hat.

Psychologische Feldexperimente scheinen besonders auch dann angebracht zu sein, wenn bei der abhängigen Variable ein konkretes, im Feld leicht beobachtbares Verhalten gemessen wird, das unter Umständen im Gegensatz zu entsprechenden verbalen Verlautbarungen steht; dies belegen zum Beispiel Studien zum Ladendiebstahl, zur Ehrlichkeit von Findern oder zu Vorurteilen. Feldexperimente eignen sich weniger zur Erfassung individueller Meinungen.

Eine inzwischen populär gewordene Vorgehensweise stellt die „Lost Letter-Technik“ dar, die erstmals von MERRITT und FOWLER verwendet und später durch MILGRAM bekannt wurde.<sup>13</sup>

11 Vgl. J. van LAWICK-GOODALL: *In the Shadow of Man*. London 1971.

12 L. A. HAYDUK: Personal space: Where we now stand. *Psychological Bulletin* 94 (1983), S. 293–335.

13 S. MILGRAM: The lost-letter technique. *Psychology Today* 3 (1969), S. 30–33, 66–68.



Bei dieser Technik werden im allgemeinen frankierte, nicht abgestempelte Briefe und Postkarten einzeln unauffällig irgendwo auf den Bürgersteig gelegt, um zu sehen, wie groß der Prozentsatz der „verlorenen“ Briefe ist, der von einem Finder aufgehoben und per Post zum Adressaten weitergeschickt wird. Die Briefe können ein Geldstück enthalten, und/oder man variiert die Adressen, den Ort und die Umstände des Verlierens usw.

Die bisherigen Erfahrungen zeigen weiter, daß trotz der Feldsituation die Bedingungen für eine logische Beweisführung, wenn auch mit Schwierigkeiten, durchaus hergestellt werden können, das heißt, Kontrolltechniken sind anwendbar, die Wirkung der unabhängigen Variable ist unter Umständen sogar besser erfaßbar.<sup>14</sup>

Die Diskrepanz zwischen Ergebnissen des Feldexperiments und vergleichbarer Laborexperimente hat in einigen Fällen auf die alarmierend geringe äußere Gültigkeit (*externe Validität*) sozialpsychologischer Laborforschung aufmerksam gemacht. Dies trifft besonders dann zu, wenn es um sozial unerwünschtes Verhalten, wie Unehrllichkeit, oder auch Einstellungen, wie Vorurteile gegenüber Minoritäten, geht.

Die Vielfalt von Vorgehensweisen, mit denen wir die Bedingungen des Verhaltens erkunden wollen – seien es experimentelle Verfahren, Feldbeobachtungen oder auch Längsschnittuntersuchungen, in denen die Bedingungen der Reifung wirksam werden –, kann hier nur angedeutet werden. Angesichts der Vielfalt erscheint es auch müßig, nach den „besten“ Methoden zu fragen. Wichtiger ist vielmehr, daß man sich der verschiedenen Möglichkeiten bewußt ist, die eigenen Forschungsprobleme gezielt einer systematischen Untersuchung zugänglich zu machen.

### Die Erfassung des Verhaltens

#### 24.2.3.

Wir würden erhebliche Schwierigkeiten haben, „das Verhalten“ umfassend und vollständig zu beschreiben. Es ist das, was wir durch psychobiologische Untersuchungen, durch Beobachtung, Experiment, Befragung, zu erklären versuchen, das Gesamt von menschlicher und tierlicher Aktivität (vgl. STE 1). Wiederum hat sich in einer konkreten Untersuchung der Forscher zu entscheiden, welcher Ausschnitt aus dem Gesamtverhalten für seine Fragestellung von Bedeutung ist.

#### Entscheidung 3

Der Anspruch, das gesamte Verhalten zu untersuchen, wäre unsinnig. Aus der Tatsache, daß der Forscher hier auswählt, wird gelegentlich der Vorwurf abgeleitet, die Psychologie als Wissenschaft betrachte den Menschen nicht als Ganzes und könne insofern nur Zerrbilder erzeugen. Zutreffend daran ist sicherlich, daß eine wissenschaftliche Untersuchung immer nur einen Teil der Wirklichkeit und damit auch des Menschen abbilden kann, und dies auch nicht anders will. Sehr wohl versucht sie, dieses Bild möglichst umfassend zu gestalten, nennt aber ausdrücklich die Möglichkeiten und die Grenzen, innerhalb derer sie sich bewegt.

*Auf welcher Ebene soll das interessierende Verhalten untersucht bzw. erfaßt werden?*

Eines der Ziele der Ethologie besteht darin, das Verhalten verschiedenster Tierarten detailliert zu dokumentieren. So sind vor allem in der Anfangsphase der Wissenschaft umfangreiche Ethogramme erstellt worden, welche die einzelnen Verhaltensweisen einer Art beschreiben und nach Funktionskreisen ordnen. Statt „Wie sieht ein bestimmtes Verhalten aus?“ wurde aber zunehmend die Frage gestellt: „Unter welchen Bedingungen tritt ein Verhalten wie häufig auf?“ Statt im Teil „Resultate“ einer Studie werden Verhaltenskategorien heute oft im Teil „Methode“ genannt.

<sup>14</sup> G. B. TUNNELL: Three dimensions of naturalness: An expanded definition of field research. *Psychological Bulletin* 84 (1977), S. 426–437.

Menschliches Verhalten läßt sich grundsätzlich auf drei Ebenen untersuchen: der physiologischen Ebene, der subjektiven Ebene und der Verhaltens-Ebene („behaviorale“ Ebene).<sup>15</sup> Da die physiologischen Methoden ein eigenes, sehr umfangreiches Forschungsfeld sind, sollen hier nur die subjektive und die Verhaltens-Ebene erörtert werden.<sup>16</sup>

Lassen sich biologische und psychologische Methoden auf der Verhaltens-Ebene noch vergleichen – die Verfahren der *Verhaltensbeobachtung* –, so ist dies für die subjektiven Informationen nicht mehr möglich. Die Psychologie genießt hier den Vorzug, über die *Befragung* des Menschen Zugang zu seinen psychischen Vorgängen zu gewinnen. Die Psychologie hat also in den subjektiven Angaben, etwa aus Interviews, Fragebogen usw., eine Informationsquelle, die dem Biologen verschlossen ist. Daß diese Informationsquelle auch ihre problematischen Seiten hat und keinesfalls frei von Verfälschungen ist, erscheint naheliegend. Verfälschungen von Verhaltensdaten können sich aber auch bei der Verhaltensbeobachtung ergeben, wenn beispielsweise das Individuum bemerkt, daß es beobachtet wird, und sich anders zu verhalten beginnt. Soll ein Verhalten ungestört erfaßt werden, muß darauf geachtet werden, daß es nicht durch Untersucher, Untersuchungssituation oder den Untersuchten verfälscht wird. Eine gewisse Hilfe können hier indirekte bzw. unauffällige Meß- und Untersuchungsverfahren bieten, die *nichtreaktiven Verfahren*, auf die wir ebenfalls in diesem Abschnitt eingehen wollen.

Vereinfacht lassen sich also drei Klassen von Untersuchungsverfahren einsetzen, um Verhalten zu erfassen:

1. Verfahren zur Beobachtung und Beschreibung des Verhaltens
2. Nichtreaktive, indirekte Meßverfahren
3. Befragungs-Verfahren (Fragebogen und Interviews).

Für tierliches Verhalten müssen wir selbstverständlich auf die unter Punkt 3 genannten Verfahren verzichten.

#### **24.2.3.1.** *Verhaltens-Beobachtung und Verhaltens-Beschreibung*

Verfahren 1

Das spezielle methodische Problem der Verhaltensbeobachtung und Verhaltensbeschreibung liegt darin, Aktivitätsmuster in Raum und Zeit festzuhalten, zu kategorisieren und in einer neuen Form darzustellen, um einerseits die Beobachtungen wissenschaftlich zu verwerten und andererseits für Dritte mitteilbar zu machen. Je nach Fragestellung muß das Verhalten auf sehr unterschiedlichen Ebenen der Komplexität beobachtet werden. Jeder Verhaltenskomplex kann in zahlreiche Unterkategorien eingeteilt werden.

So läßt sich zum Beispiel der Kampf zweier Männchen um ein Territorium unterteilen in: „droht“, „springt Gegner an“, „beißt“, „kratzt“ usw. Dazu könnte man festhalten, auf welche Körperteile des Gegners das jeweilige Verhalten gerichtet ist. Es wäre jedoch auch möglich, einfach den Gewinner und Verlierer festzustellen: dieses Vorgehen würde etwa für die Frage genügen, ob ein Eindringling in einem Kampf die gleichen Chancen hat wie ein Territoriumsbesitzer. Möchte man jedoch wissen, ob der Kampf zwischen zwei etwa gleich großen Gegnern eher eskaliert als zwischen zwei verschiedenen großen, müßten mehrere Teilelemente erfaßt werden, welche die Intensität und Härte eines Kampfes ausdrücken. In gewissen Fällen schließlich beobachtet ein Untersucher gar keine einzelnen Verhaltensweisen, sondern nur deren Auswirkung. Ein Beispiel dafür war das Auslegen von Eiattrappen in der Nähe von brütenden Möwen. Die Autoren versuchten nicht zu ermitteln, in welcher Weise die Brutvögel die Attrappen gegenüber Raubfeinden schützten, sondern stellten lediglich fest, welcher Prozentsatz der Attrappen unversehrt blieb.<sup>17</sup>

15 N. BIRBAUMER: Psychophysiologie der Angst. München 1977.

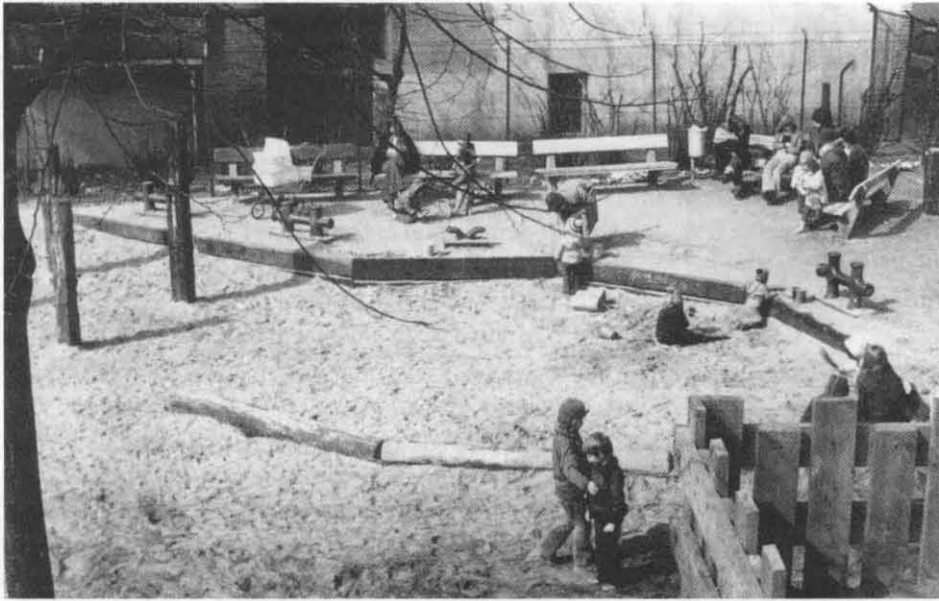
16 Vgl. R. SCHANDRY: Psychophysiologie – Körperliche Indikatoren menschlichen Verhaltens. München 1981.

17 F. GÖTMARK / M. ANDERSSON (s. *Anm.* 3).

Für die Biologie, speziell die Ethologie, wie auch für die Psychologie liefert die systematische Beobachtung wesentliche Informationen über das Verhalten. Vielfältig sind die entsprechenden Methoden, von denen hier nur einige Aspekte dargestellt werden können. Für die gründliche Auseinandersetzung mit diesem Gebiet muß auf Übersichten verwiesen werden.<sup>18</sup>

Am Beispiel einer Verhaltensbeobachtung von Kindern im Sandkasten (Abb. 8) lassen sich verschiedene Beobachtungsverfahren und deren Komponenten erläutern.

Abb. 8: Beobachtungsfeld Kinderspielplatz



Aus: Kinderspielplätze. Beiträge zur kindorientierten Gestaltung der Wohnumwelt. Teil 2: Bildband. Schriftenreihe des Bundesministers für Jugend, Familie und Gesundheit. Bd. 44/2. Stuttgart <sup>2</sup>1977, S. 153.

Auf die Segmentierung des Verhaltens, auf Interpretations-Aspekte, auf einige Merkmale von Beobachtungssystemen und auf die Güte-Kriterien für die Beobachtung wollen wir näher eingehen:

**Segmentierung:** Dem Beobachter auf der Parkbank bietet sich bei den spielenden Kindern ein kontinuierlicher Verhaltensstrom dar. Versucht er, diesen Verhaltensstrom zu beschreiben, so wird er zunächst bemerken, daß im Spiel verschiedene voneinander abgrenzbare Ereignisse auftreten. Diese Ereignisse lassen sich als Verhaltenseinheiten oder Episoden festhalten. Diese „Segmentierung“ des Verhaltensstroms scheint zunächst den Ablauf in „natürliche Einheiten“ zu trennen: Erst spielt das Kind, dann läuft es zur Mutter, dann wieder zurück zu den anderen Kindern usw. Aber schon bei näherer Betrachtung des Spiels wird man erkennen, daß dieser Ablauf selbst wieder verschiedene Episoden enthält, die ineinander übergehen: Das Kind beginnt, allein einen Sandkuchen zu formen; ein anderes Kind kommt hinzu; beide formen gemeinsam; eines beginnt, die Kuchen zu zerstören, usw. Dies Spiel gliedert sich wiederum in einen Ablauf mit verschiedenen Ereignis-

18 Vgl. M. VON CRANACH / H. G. FRENZ: Systematische Beobachtung. In: C. F. GRAUMANN (Hrsg.): Handbuch der Psychologie. Bd. 7: Sozialpsychologie, 1. Halbband: Theorien und Methoden. Göttingen 1969, S. 269–331. – U. MEES / H. SELG: Verhaltensbeobachtung und Verhaltensmodifikation. Stuttgart 1977. – K. R. SCHERER / P. EKMAN (Hrsg.): Handbook of Methods in Nonverbal Behavior Research. Cambridge 1982. – J. ALTMANN: Observational study of behavior: Sampling methods. *Behaviour* 49 (1974), S. 227–267.

sen. In eindeutiger Weise lassen sich also „natürliche“ Einheiten nur schwer definieren. Selbst für unsere Sprache, die doch klar durch die Grammatik gegliedert scheint, bereitet eine eindeutige Segmentierung Schwierigkeiten. Eine gesprochene Äußerung kann zum Beispiel aus einem einzigen „Nein“ bestehen, aus einem unvollständigen oder aus einem längeren Satz. Die Zusammenfassung nach „Sinn-einheiten“ liegt nahe – doch auch hierfür ergeben sich Schwierigkeiten. Das Problem der Segmentierung ist insofern von Bedeutung, als sich hier entscheidet, mit welchen quantitativen Daten man später den beobachteten Ablauf kennzeichnet, welche Häufigkeiten und Dauern von Ereignissen man feststellt.

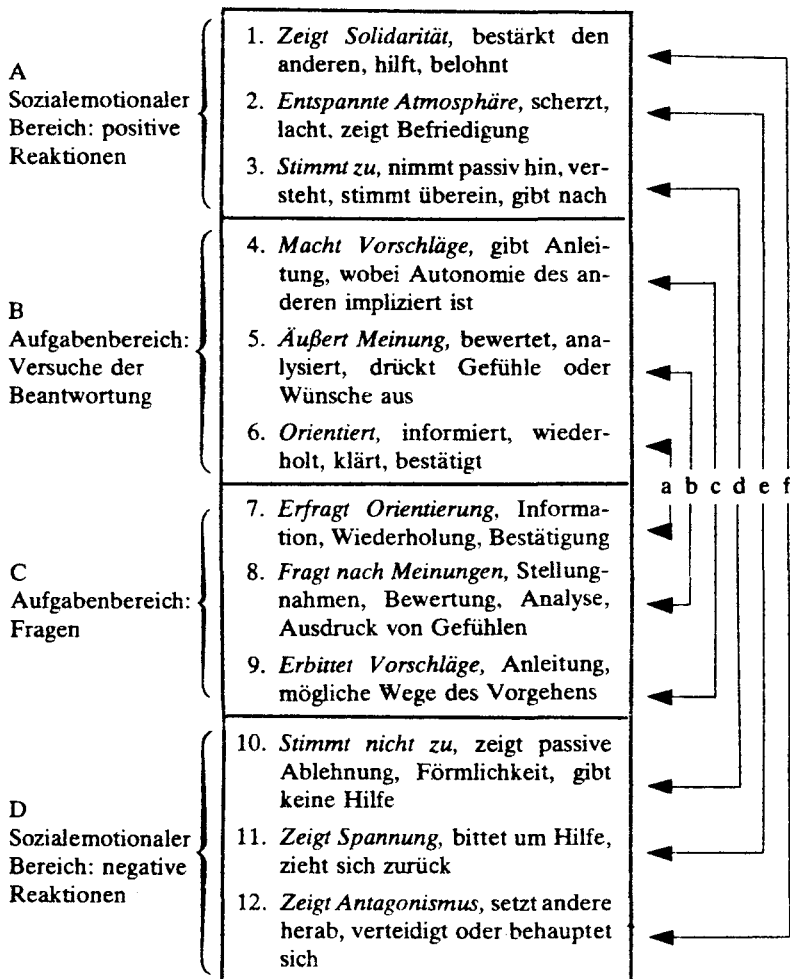
*Quantifizierung:* Eine „Quantifizierung“ von Verhaltens-Informationen wiederum ist notwendig, um verschiedene Situationen vergleichen zu können. Die Frage, ob Kinder durch die Anwesenheit eines anderen Kindes zu längerer oder intensiverer Beschäftigung mit dem Spiel angeregt werden, kann man nur über irgendeine Form des quantitativen Vergleichs beantworten. Auch für qualitative Beobachtungen wird, wenn auch teils unausgesprochen, ein quantitativer Vergleich getroffen, sofern sich die Aussage auf „mehr“ oder „weniger“ eines Merkmals bezieht. Stellt man beispielsweise fest, daß Kinder gemeinsam ein „kreativeres“ Spiel gestalten als allein, so ist darin eine quantitative Aussage enthalten. Auf einer gedachten Skala stuft man das Verhalten als mehr oder weniger „kreativ“ ein.

*Operationalisierung:* Die Definition dessen, was man unter einer Verhaltenskategorie fassen will, stellt einen schwierigen Teil der Verhaltensbeobachtung dar. Welches Verhalten ist als „kreatives Spiel“ zu bewerten? Ist es die Vielfalt hergestellter Sandformen, die Verwendung von zusätzlichen Materialien, wie Holzstöckchen, Wasser usw.? Das, was uns im Alltagsverständnis als klarer Begriff erscheint, wird dann problematisch, wenn man genauer angeben soll, welche Verhaltensaspekte zu einer Kategorie gehören und welche nicht. Ein Weg, um zu einer vergleichbaren Verwendung von Begriffen und Kategorien zu gelangen, ist deren „Operationalisierung“. Man spricht von „kreativ“ und gibt gleichzeitig die „Operationen“ an, die dieses Konzept definieren. Auch am Beispiel der Aggression läßt sich erkennen, wie wichtig solche operationalisierten Definitionen für eine wissenschaftliche Verständigung sind. So sollte man für die Beobachtung vor der Untersuchung angeben, welche Verhaltensweisen als aggressiv zu werten sind – ob es sowohl sprachliche als auch nichtsprachliche Verhaltensweisen sein sollen, ob bereits ein negativer Gesichtsausdruck ausreicht, usw. Nur dann kann man sicherstellen, daß verschiedene Beobachter und Untersucher ein ähnliches Verständnis von dem haben, was sie als „Aggression“ bezeichnen. Die operationale Definition der Verhaltenskategorien ist besonders kritisch bei globalen Kategorien, wie bei den hier genannten Merkmalen „Kreativität“ oder „Aggressivität“. Daher versucht man, für Beobachtungssysteme solche Kategorien zu entwickeln, die möglichst eindeutige Beschreibungen auf der Verhaltensebene beinhalten.

*Beobachtung und Interpretation:* Ziel einer systematischen Beobachtung ist die möglichst objektive Beschreibung dessen, was beobachtet wurde. Das bedeutet für den Beobachter, daß er selbst weitgehend vermeiden sollte, das Verhalten zu interpretieren. Im Alltag sind wir dagegen gewohnt, Verhalten gleichzeitig zu beobachten und zu interpretieren. Man gibt dem Verhalten einen Sinn, indem man etwa dem Handelnden bestimmte Motive unterstellt, die sein Verhalten für uns erklärbar machen. Eine Feststellung wie die, daß ein Kind dem anderen aus Ärger oder Eifersucht die geformten Sandkuchen zerstört, beinhaltet Interpretationen, die kaum nachprüfbar sind: Es wird ein Motiv unterstellt, etwa dem anderen zu schaden, ihn zu ärgern, vielleicht auch ein Versuch, dem anderen gegenüber dominant zu sein.

Allerdings enthalten auch verschiedene Beobachtungs-Systeme interpretative Anteile. Verhaltenskategorien aus einem Beobachtungssystem von BALES (Tab. 1) zur Erfassung von Gruppen-Interaktionen und dem Verhalten einzelner Personen in Gruppen sollen als Beispiel für eine mittlere Interpretationsebene dienen:<sup>19</sup>

Tab. 1: Kategorien-System nach Bales



**Schlüssel:**

- |                               |                                       |
|-------------------------------|---------------------------------------|
| a – Probleme der Orientierung | d – Probleme der Entscheidung         |
| b – Probleme der Bewertung    | e – Probleme der Spannungsbewältigung |
| c – Probleme der Kontrolle    | f – Probleme der Integration          |

Aus: G. FASSNACHT: Systematische Verhaltensbeobachtung. Eine Einführung in die Methodologie und Praxis. München 1979, S. 114.

Für den sozial-emotionalen Bereich umschreiben die folgenden Kategorien positive Reaktionen, nach denen das Verhalten einer Person kodiert bzw. klassifiziert werden kann:

- [1] *Zeigt Solidarität*, bestärkt den anderen, hilft, belohnt;
- [2] *Entspannte Atmosphäre*, scherzt, lacht, zeigt Befriedigung;
- [3] *Stimmt zu*, nimmt passiv hin, versteht, stimmt überein, gibt nach.

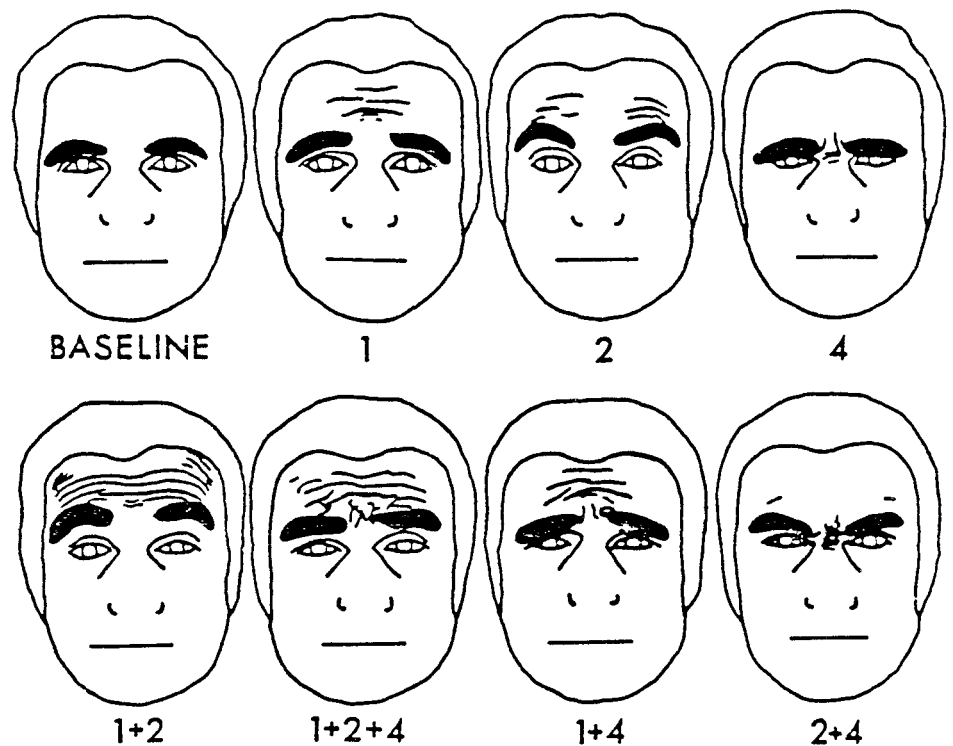
<sup>19</sup> Vgl. R. F. BALES: Die Interaktionsanalyse. Ein Beobachtungsverfahren zur Untersuchung kleiner Gruppen. In: R. KÖNIG (Hrsg.): Beobachtung und Experiment in der Sozialforschung. Köln <sup>8</sup>1972, S. 148–167.

Es werden hier jeweils eine Reihe von unterschiedlichen Verhaltensaspekten in einer Kategorie zusammengefaßt. In der Kategorie [2] sind es zum Beispiel verbale und nonverbale Elemente. Der Beobachter ist gleichzeitig aufgefordert, „Befriedigung“ aus dem Verhalten zu interpretieren.

Interessiert man sich genauer für das Gebiet der emotionalen Zustände, so wird solch eine globale Bewertung nicht ausreichen. Man wird hingegen dem Ausdrucksverhalten mehr Aufmerksamkeit schenken. Die Stimme und die Mimik scheinen hierfür besonders aufschlußreiche Verhaltensweisen zu sein. Stimmliche Merkmale lassen sich durch Messungen der Frequenzen, vor allem der Grundfrequenz, und der Lautstärke objektiv erfassen und vergleichen.<sup>20</sup>

Für die Analyse der Mimik wurde von EKMAN und FRIESEN ein Beobachtungssystem auf der Grundlage eines Systems des schwedischen Astronomen HJORTSJÖ entwickelt, das es erlaubt, die Bewegungen im Gesicht interpretationsfrei zu beschreiben.<sup>21</sup> Für die Stirn bilden 3 Elemente, die einzeln oder in Kombination auftreten, das Repertoire für Verhaltenseinheiten (*Action Units*) (Abb. 9).

Abb. 9: Elemente und Beobachtungs-Codes der Mimik im Stirnbereich



Nach: P. EKMAN: About brows: Emotional and conversational signals. In: M. von CRANACH / K. FOPPA / W. LEPENIES / D. PLOOG (Hrsg.): Human Ethology. London 1979, S. 174.

Solche Action Units laufen zum Teil sehr rasch (kürzer als 1 Sekunde) ab und lassen sich häufig nur im Zeitlupentempo vom Videoband genauer bestimmen. Wendet man dieses System an, so erfolgt erst nach der Beobachtung und Registrierung in einem zweiten Schritt die Interpretation des Verhaltens. Man stellt fest, welcher emotionale Ausdruck vorhanden war, ob das Verhalten als reine oder gemischte Darstellung einer Emotion oder als „Konversationssignal“ zu betrachten ist.

<sup>20</sup> K. R. SCHERER: Beobachtungsverfahren zur Mikroanalyse nonverbaler Verhaltensweisen. In: G. van KOOLWIJK / M. WIEKEN-MAYSER (Hrsg.): Techniken der empirischen Sozialforschung. Bd. 3. München 1974, S. 66–109. – Ders.: Ausgewählte Methoden der empirischen Sprachforschung. In: Ebda., S. 110–157.

<sup>21</sup> P. EKMAN / W. V. FRIESEN: Facial Acting Coding System. Palo Alto, Cal. 1978.

Wenn hier Kategorien des Systems von BALES und des „Facial Action Coding Systems“ von EKMAN und FRIESEN gegenübergestellt wurden, so sollten damit vor allem die unterschiedlichen Interpretationsebenen illustriert werden, auf denen Verhalten beobachtet wird. Auch hier gilt, daß die Wahl der Beobachtungskategorien und der darin enthaltenen Interpretationsanteile abhängig von der jeweiligen Fragestellung ist.

### *Gütekriterien der Beobachtung*

Geben wir zwei Spielplatz-Beobachtern die Aufgabe, zwei dort miteinander spielende Kinder zu beobachten und zu erfassen, wie sehr sie miteinander kooperieren, so findet der eine, daß sie viel gemeinsam machen, einander zur Hand gehen und sich besonders oft ansehen, den anderen beeindrucken vielleicht ihre stimmlichen und sprachlichen Äußerungen, ihr freundlicher Tonfall und ihre Absprachen. Beide achten also auf verschiedene Dinge, und es erscheint schwierig, zu vergleichbaren Aussagen zu gelangen. Die Beobachter würden zu zwei unterschiedlichen Meßwerten, beispielsweise zwei differierenden Punktwerten für „Kooperation“, gelangen, und die Messung bzw. Beobachtung wäre insgesamt nicht sehr zuverlässig (*reliabel*).

Weiterhin stellt sich die Frage, ob eine Beobachtung überhaupt das erfaßt, was sie messen will. Sagen die von unseren Beobachtern verwendeten Verhaltenskategorien (z. B. „etwas gemeinsam machen“, „Absprachen“) überhaupt etwas über „kooperatives Verhalten“ aus, und wie gut bzw. genau (*valide*)? Wir müssen also nach der Güte einer Beobachtung fragen.

**Die Reliabilität:** Die „Zuverlässigkeit“ ist eines der wichtigsten Kriterien, nach denen die Güte einer Beobachtung zu überprüfen ist.

Gibt man beispielsweise zwei Beobachtern die Aufgabe, die drei oben genannten Kategorien zum sozial-emotionalen Bereich im Kategoriensystem von BALES unabhängig voneinander zu registrieren, so wird man unmittelbar feststellen, daß nur zu einem gewissen Prozentsatz Übereinstimmung zwischen den Beobachtern besteht. Auch die Feststellung, ob eine Person die Action Units 4 oder die Action Units 1 + 4 gezeigt hat, werden zwei unabhängig mit dem „Facial Action Coding System“ arbeitende Beobachter nicht immer übereinstimmend treffen.

Das Ausmaß der Übereinstimmung zweier Beobachter wird als „Inter-Rater-Reliabilität“ bezeichnet. Die Reliabilität der Beobachtung, angegeben als prozentuale Übereinstimmung oder als Korrelations-Koeffizient (Kap. 24.2.4) zwischen zwei unabhängig arbeitenden Beobachtern, ist eine notwendige Angabe über die Verlässlichkeit von Daten. Eine hohe Reliabilität wäre zum Beispiel bei 85% bis 90% Übereinstimmung gegeben (wobei Zufallskorrekturen zu berücksichtigen sind).<sup>22</sup> Eine geringe Beobachterübereinstimmung kann auf verschiedenen Ursachen beruhen. Sie kann ein Hinweis darauf sein, daß das Beobachtungssystem ungenügend definierte Kategorien enthält, daß die Beobachter nicht hinreichend geschult wurden oder daß sie unterschiedliche, unausgesprochene Konzepte der zu beobachtenden Phänomene hatten. Erst eine Überprüfung der Reliabilität erlaubt eine gezielte Verbesserung des Beobachtungs-Instruments oder ein verbessertes Training der Beobachter.

**Die Validität:** „Gültigkeit“ verlangt von einem Beobachtungsinstrument, daß es das in Frage stehende Phänomen auch tatsächlich gültig (*valide*) erfaßt. Wir verstehen unter „Validität“ den Grad der Genauigkeit, mit dem eine Beobachtung diejenige Verhaltensweise oder dasjenige Persönlichkeitsmerkmal tatsächlich mißt, die/das sie messen soll oder zu messen vorgibt. Das Problem der Validität der Beobachtungen stellt sich bei jeder Interpretation von Ergebnissen, die durch systematische Beobachtungsverfahren gewonnen wurden.

<sup>22</sup> J. ASENDORPF / H. G. WALLBOTT: Maße der Beobachterübereinstimmung. Ein systematischer Vergleich. *Zeitschrift für Sozialpsychologie* 10 (1979), S. 243–252.

Zu fragen ist beispielsweise, ob die Kategorien des Systems von BALES die angenommenen Vorgänge auch tatsächlich erfassen. Oder welche der von EKMAN und FRIESEN untergliederten mimischen Elemente den Ausdruck bestimmter Emotionen gültig bestimmen.

Insgesamt ist bei der Bewertung von systematischer Beobachtung zu bedenken, daß es sich um sensible Meßinstrumente handelt. Die Sensibilität bezieht sich dabei auch auf mögliche Einflüsse des Messenden bzw. des Beobachters.<sup>23</sup> Beobachtungen beinhalten hochkomplexe Verarbeitungsvorgänge beim Beobachter. Um so notwendiger ist eine Überprüfung dieser Beobachtungsleistung, um die Fehlermöglichkeiten möglichst gering zu halten.

#### *Verhaltensstichproben*

Für jede Verhaltensbeobachtung ist ein geeignetes Beobachtungsraster zu wählen. Man unterscheidet hierbei gewöhnlich zwischen „Zeit-Stichproben“ (*time sampling*) und „Ereignis-Stichproben“ (*event sampling*):

- In einer *Zeitstichprobe* werden Zeitpunkte oder Zeitabschnitte der Beobachtung festgelegt. Man kann beispielsweise festlegen, daß das Verhalten von nur 1 Kind im Sandkasten bei jeder neuen Minute registriert wird, oder man stellt beispielsweise für jeden Zeitraum von 1 Minute fest, ob es allein oder mit anderen spielt, und gewinnt so einen Überblick über die zeitliche Verteilung der Aktivität.

- Bei einer *Ereignis-Stichprobe* hält man hingegen lediglich die Tatsache fest, daß ein zuvor definiertes Verhalten aufgetreten ist. Beispiel für eine Ereignisstichprobe wäre das Registrieren von aggressivem Verhalten der Kinder im Sandkasten. Sobald man solch ein Ereignis bemerkt hat, hält man es fest. Dies kann etwa dadurch geschehen, daß man eine Strichliste anfertigt oder die Beobachtung in einem Protokoll notiert.

Die Form der Beobachtungsstichprobe – *time sampling* oder *event sampling* – muß dem jeweiligen Verhalten angemessen sein. Ein Verhalten, das sehr selten auftritt, wird man in einer Ereignisstichprobe festhalten. Es wäre beispielsweise unsinnig, mit einer Zeitstichprobe registrieren zu wollen, ob ein Kind die Spielgruppe beleidigt verläßt. Dies Verhalten tritt vielleicht innerhalb des Beobachtungszeitraums nur einmal auf. Durch eine Zeitstichprobe würde dieses wichtige Ereignis durch die Menge unwichtiger registrierter Informationen überdeckt. Andererseits wird man in solchen Fällen, in denen es auf die Dauer des Verhaltens ankommt, eine Zeitstichprobe wählen. Wenn man beispielsweise daran interessiert ist, wie sich das Spielverhalten im Verlaufe der Entwicklung vom 3. zum 5. Lebensjahr ändert, so könnte man als Bezugsgröße die Zeitdauer bestimmen, mit der sich Kinder verschiedener Altersstufen unter vergleichbaren Bedingungen im gemeinsamen Spiel betätigen. Hier läge die relevante Information nicht im Einzelereignis, sondern in der zeitlichen Erstreckung und der Verteilung über einen Spielvormittag.

#### *Praktische Aspekte der Verhaltenserfassung*

Bleistift, Papier und Stoppuhr gehören seit frühester Zeit zur Ausrüstung eines jeden Ethologen und Psychologen. Wichtigste technische Hilfsmittel der „höheren Art“ sind jedoch Tonband, Fotoapparat und Film- oder Videokamera. Spezielle Hochfrequenz-Kameras erlauben eine Analyse extrem schneller Bewegungsabläufe. Je nach Größe und Lebensart der Tiere sind oft Fernglas und Spektiv oder aber Lupe und Mikroskop notwendig.

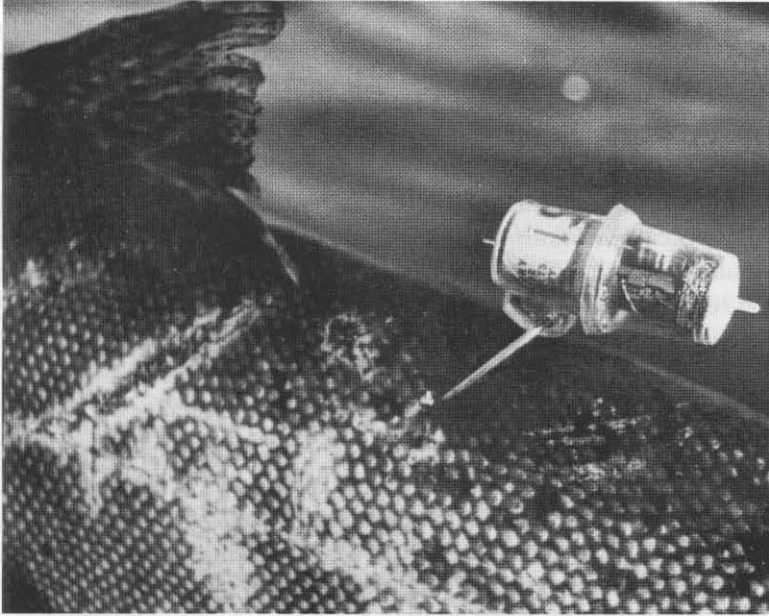
<sup>23</sup> Vgl. D. T. CAMPBELL: Systematic error on the part of human links in communication systems. *Information and control* 1 (1958), S. 334–369.



Große Bedeutung hat auch der Ereignisrekorder, ein mehr oder weniger kleines Kästchen mit einer Tastatur. Das Innere des Kästchens zeichnet auf, wie lange jeweils welche Taste gedrückt wird. Modernere Geräte speichern diese Information elektronisch und erlauben, diese an einen Computer zu übermitteln. Ein Untersucher kann nun jeder Taste ein bestimmtes Verhalten oder Tier zuordnen und – mit genügend Übung – seine Aufzeichnungen machen, ohne die Augen von den Tieren abzuwenden.

Um Aktivitätsmuster von Tieren aufzunehmen, die entweder sehr große Distanzen zurücklegen oder ein verborgenes Leben führen, werden häufig Radiosender verwendet. Sie dienen in erster Linie dazu, die jeweilige Position des Trägers mit Hilfe eines Empfängers zu orten. Verbunden mit Meßgeräten, können sie aber auch Informationen über den inneren Zustand des Tieres, zum Beispiel die Körpertemperatur, übermitteln (*Abb. 10*).

*Abb. 10:* Praktische Verhaltenserfassung



Um das Orientierungsvermögen von Tieren zu erforschen, werden heute kleine Transistorgeräte benutzt, die z. B. die Atemtätigkeit, den Herzschlag oder Kursänderungen registrieren und über einen Kleinstsender an weit entfernte Empfangsstationen übermitteln. Hier: Lachs mit einem der beschriebenen Sender.

Aus: N. TINBERGEN und die Redaktion der Time-Life-Bücher: Tiere und ihr Verhalten. Reinbek 1976, S. 109.

Für viele Verhaltensstudien ist es unabdingbar, die einzelnen Tiere individuell unterscheiden zu können. Dies ist dann verhältnismäßig einfach, wenn die Tiere ein auffälliges, individuelles Muster auf ihrer Haut tragen, wie etwa die Streifen der Zebras. Mit entsprechender Übung kann ein Mensch jedoch auch wesentlich feinere Unterschiede wahrnehmen und so beispielsweise viele Dutzend Affen innerhalb einer Gruppe anhand von Gesichtszügen, Haltung, Gang usw. unterscheiden. In vielen Fällen müssen die Tiere jedoch individuell markiert werden. Dies ist eine Art Wissenschaft für sich, und jede Tiergruppe erfordert besondere Verfahren. Eine umgehängte, angeklebte oder aufgemalte Etikette sollte das Tier weder behindern noch zu einem Außenseiter für Artgenossen machen. Im Freiland ist darauf zu achten, daß die Markierung das Tier nicht gegenüber Räubern hervorhebt.

### 24.2.3.2. *Der Einsatz nichtreaktiver Meßverfahren*

#### Verfahren 2

Besonders bei der Untersuchung menschlichen Verhaltens ist damit zu rechnen, daß die Untersuchung selbst einen Einfluß auf das interessierende Phänomen hat und daß solche Effekte schwer zu kontrollieren sind. Ein möglicher Ausweg aus diesem methodischen Dilemma besteht nun naheliegenderweise darin, daß empirische Untersuchungen so durchgeführt werden, daß die Betroffenen gar nicht wahrnehmen, daß sie „Objekt“ einer wissenschaftlichen Studie sind. Man nennt solche Verfahren „nicht-reaktive“ bzw. „unauffällige“ Verfahren. Die Untersuchung mit derartigen nicht-reaktiven Meßinstrumenten findet natürlich nicht im Labor, sondern in der Regel in einer natürlichen Umgebung statt, und der Forscher oder Versuchsleiter ist als solcher nicht erkennbar.

In ihrer inzwischen schon „klassischen“ und nach wie vor höchst aktuellen Publikation aus dem Jahre 1966 haben WEBB, CAMPBELL, SCHWARTZ und SECHREST eine Fülle derartiger Forschungsmöglichkeiten zusammengestellt.<sup>24</sup>

Der Kanon der präsentierten Verfahren umfaßt dabei durchaus neuartige, originelle und zum Teil amüsante Ideen; er enthält aber auch gleichzeitig traditionelle, altbewährte, beinahe schon vergessene Techniken der Sozialwissenschaften, die im Lichte der Reaktivitätsproblematik (Kap. 24.4.1) eine Renaissance erleben:

#### *Analyse physischer Spuren: Abnutzungen und Ablagerungen*

Die Autoren berichten beispielsweise von den Beobachtungen einer Kommission, daß der Fußboden rund um die Glaskästen mit schlüpfenden Kücken in *Chicago's Museum of Science and Industry* etwa alle sechs Wochen erneuert werden mußte, während andere Teile des Fußbodenbelags nur alle paar Jahre reparaturbedürftig waren. Wie zusätzliche Beobachtungsdaten belegen, zeugen diese Abnutzungen für überdurchschnittliches Interesse der Besucher. Ähnliche Verschleißerscheinungen, etwa bei Büchern in einer Bibliothek, könnten somit ökonomisch und effektiv für nichtreaktive „Beliebtheitsermittlungen“ Verwendung finden.

Ebenso können Ablagerungen im weitesten Sinne, also zum Beispiel Abfälle, näher untersucht werden. Leere Bier- und Schnapsflaschen, leere Zigarettenschachteln und ähnliche Dinge bleiben nach Fußballspielen, Karnevalszügen und Popkonzerten zurück und werden auch tatsächlich von Marktforschungsinstituten zur Berechnung von Marktanteilen einzelner Produkte ausgezählt.<sup>25</sup>

Stärker experimentell orientierte Untersuchungen zum Leserverhalten haben versucht, das Interesse an Anzeigen in Zeitschriften zu ermitteln, indem sie die jeweiligen Druckseiten chemisch so präparierten, daß man im nachhinein feststellen konnte, ob sie vom Leser berührt wurden.

Schließlich kann in diesem Zusammenhang eine Studie erwähnt werden, bei der ebenfalls Spuren im weitesten Sinne, nämlich Beschriftungen (Grafitti) auf Toiletten, ausgewertet wurden und signifikante Unterschiede zwischen Herren- und Damentoiletten ermittelt werden konnten.<sup>26</sup> Solche Techniken können im übrigen auch im Tierbereich eingesetzt werden. Bekannt ist beispielsweise, daß sich das Beuteverhalten der Fische sehr leicht aus ihren „Abfällen“ vor dem Bau erschließen läßt.

24 E. H. WEBB / D. T. CAMPBELL / R. D. SCHWARTZ / L. SECHREST: Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago 1966 (deutsch: Nichtreaktive Meßverfahren. Weinheim/Basel 1975).

25 H. E. LÜCK / U. REUTER: Marktforschung kann mehr leisten. *Marketing Journal* 1/1973, S. 72–74, 2/1973, S. 162–165. – C. B. MERRITT / R. G. FOWLER: The pecuniary honesty of the public at large. *Journal of Abnormal and Social Psychology* 43 (1948), S. 90–93.

26 A. C. KINSEY / W. B. POMEROY / C. E. MARTIN / P. H. GEBHARD: Sexual Behavior in the Human Female. Philadelphia 1953.

### Inhaltsanalyse archivarischer Daten

Die Möglichkeit einer Analyse archivarischer Materials wird von WEBB und Kollegen als besonders geeignete Form einer nichtreaktiven Messung dargestellt und diskutiert. Personenstandsregister können zur Untersuchung vorehelicher Geschlechtsbeziehungen in verschiedenen Gesellschaften<sup>27</sup> oder auch des Zusammenhangs zwischen religiöser Identifizierung und beruflichem Status verwendet werden.<sup>28</sup> Berühmt sind in diesem Zusammenhang auch die Studien von DURKHEIM über den Selbstmord.<sup>29</sup> Er fand Häufungen zu bestimmten Jahreszeiten und Wochentagen. Politische und Gerichts-Berichte, also etwa Wahlstatistiken, politische Reden, Gerichtsurteile, Polizeiberichte, oder sonstige öffentliche Berichte, über Energieverbrauch, Benutzung von Parkuren oder periodische Schwankungen des Wasserdrucks, liefern ebenfalls eine Fülle interessanter Materials, das sozialwissenschaftlich genutzt werden kann.<sup>30</sup>

Leicht zugänglich sind des weiteren die kontinuierlichen Daten, die sich den Massenmedien entnehmen lassen, wie die einschlägige Literatur über Inhaltsanalysen belegt.<sup>31</sup> DOLLASE überprüfte beispielsweise Vorurteile anhand der Liederbewertungen beim „Grand Prix de la Chanson“ in den Jahren 1968 bis 1972 mit Hilfe soziometrischer Techniken, Verfahren, die soziale bzw. zwischenmenschliche Beziehungen messen.<sup>32</sup> Aber auch einzelne episodische und private Berichte im weitesten Sinne können die Ausgangsdaten für nichtreaktive Auswertungsverfahren darstellen. Man kann sich etwa Verkaufsberichte von Gesellschaften ansehen, um an dem Verkauf von alkoholischen Getränken die durch Flugunfälle hervorgerufene Furcht zu messen, um an den Umsatzzahlen einzelner Medikamente die Informationsverbreitung unter Ärzten abzulesen oder auch nur, um die Effizienz einer Werbekampagne zu testen. DÜNNBIER hat in seiner Studie über „Verbrecher und Verbrechertum in Bayern 1882 bis 1937“ einen Zusammenhang zwischen Bierkonsum und Verurteilungsziffern für Körperverletzungen festgestellt.<sup>33</sup> Im Bereich der Industriepsychologie können Industrie- und Institutionenberichte (also Kündigungsraten, Anzahl der Arbeitsplatzwechsel, Beschwerdeberichte, Krankmeldungen usw.) wertvolle Hinweise auf Motivationsstrukturen der Arbeitnehmer geben.

### Nichtreaktive Beobachtungen

Die vielleicht grundlegendste Methode der Psychologie, die in fast allen Verfahren der empirischen Datenerhebung in weiterem Sinne enthalten ist, nämlich die Beobachtung, kann ebenfalls unter die nichtreaktiven Meßverfahren gezählt werden, wenn spezifische Voraussetzungen gegeben sind. Entscheidend ist, daß die Personen nicht bemerken, daß ihr Verhalten vom Forscher beobachtet wird. Im einfachsten Fall nimmt der Forscher eine unauffällige, unaufdringliche passive Rolle ein. Er registriert äußerliche Erscheinungsmerkmale, wie Tätowierungen bei Gefängnisinsassen, die Kleidung von Kranken in ambulanter Behandlung, politische Aufkleber auf Pkws oder Ausdrucksbewegungen, um entsprechende

27 H. T. CHRISTENSEN: Cultural relativism and premarital sex norms. *American Sociological Review* 25 (1960), S. 31–39.

28 L. G. BURCHINAL / W. F. KENKEL: Religious identification and occupational status of Iowa grooms. *American Sociological Review* 27 (1962), S. 526–532.

29 E. DURKHEIM: *Le Suicide*. Paris 1897 (deutsch: *Der Selbstmord*. Neuwied 1973).

30 W. BUNGARD / H. E. LÜCK: *Forschungsartefakte und nichtreaktive Meßverfahren*. Stuttgart 1974.

31 R. LISCH / J. KRIZ: *Grundlagen und Modelle der Inhaltsanalyse*. Reinbek 1978.

32 R. DOLLASE: Beurteilung von populären Liedern im Grand Prix de la Chanson Eurovision. *Psychologie und Praxis* 17 (1973), S. 26–33.

33 H. DÜNNBIER: *Verbrechen und Verbrechertum in Bayern 1882 bis 1937*. Beiträge zur Statistik Bayerns. Bd. 138. München 1944.

Schlußfolgerungen daraus zu ziehen. Ein großer Teil der nonverbalen Kommunikationsforschung beruht auf derartigen Beobachtungsstudien.<sup>34</sup>

Sehr interessante Forschungen wurden so auch zur Analyse des Sprachverhaltens durchgeführt, indem unauffällig die Gespräche von Passanten auf der Straße oder auf Bahnhöfen festgehalten wurden, unter anderem, um geschlechtsspezifische Differenzen ausfindig zu machen.

LAUTMANN untersuchte als teilnehmender Beobachter (in der Rolle eines Gerichtsassessors) die richterlichen Strategien zum Abschluß eines Verfahrens.<sup>35</sup> Interessant ist auch die Studie von SCHUMANN und WINTER, die als „Zuschauer“ im Gerichtssaal die Interaktionsformen zwischen Richtern, Zeugen und Angeklagten systematisch beobachteten.<sup>36</sup> Mit dem zunehmenden Einsatz technischer Hilfsmittel sind die Möglichkeiten derartiger nichtreaktiver Beobachtungsstudien in den letzten Jahrzehnten ständig erweitert worden. Mit Hilfe von Ton-, Foto- (u. a. auch Infrarotfotos) und vor allem Videoaufnahmen können menschliche Verhaltensweisen unbemerkt registriert werden, die sonst unter Umständen einer direkten Beobachtung durch den Forscher nicht zugänglich gewesen wären.

Auch die Bemerkungen zur ethologischen Feldforschung weisen auf das Bemühen hin, bei der Beobachtung des Verhaltens von Tieren in ihrer natürlichen Umgebung möglichst wenig in das Verhalten einzugreifen und somit „nichtreaktiv“ vorzugehen.

Außerdem verwenden Ethologen nichtreaktive Verfahren auch dann, wenn Tiere einer Beobachtung nicht zugänglich sind.

So läßt sich beispielsweise das Freileben des Sumatranashorns ausschließlich aus der sorgfältigen Erfassung und Analyse von Spuren (Fraßspuren, Fährten, Markierungen usw.) erfassen. Die vom Aussterben bedrohten Nashörner leben in geringen Dichten im tropischen Regenwald Sumatras, und es ist äußerst schwierig, überhaupt ein Exemplar aufzufinden. Im allgemeinen wird ein Ethologe ein Tier wählen, das er genügend häufig beobachten kann; manchmal wird es jedoch notwendig, auch sehr seltene, scheue und in unübersichtlichem Gelände wohnende Arten zu studieren, insbesondere wenn dadurch ein Beitrag zu ihrem Schutz geleistet werden kann. Darüber hinaus sparen manchmal nichtreaktive Verfahren erheblich Zeit und Kosten (etwa eine mit einem elektronischen Zähler verbundene Lichtschranke vor dem Eingang eines Baues).

Trotz der offensichtlichen Vorteile von nichtreaktiven Verfahren sollte man nicht deren teils erhebliche methodische Probleme übersehen:

- Die Stichproben sind oft selektiv und bezüglich der Repräsentativität nicht überprüfbar. (Wie sind Archive entstanden? Welche Spuren halten sich besser? usw.)
- Die Gütekriterien sind nicht immer überprüfbar. Beobachtungen sind unter Umständen selektiv und damit verzerrt; bei Feldexperimenten gibt es zahlreiche unkontrollierbare Einflußfaktoren.
- Nichtreaktive Meßverfahren können doch reaktiv werden, wenn die Techniken bekanntwerden (verlorene Briefe) oder wenn der unauffällige Beobachter „auffällig“ wird.
- Die Techniken sind inhaltlich begrenzt, das heißt, sie sind nur auf bestimmte Fragestellungen anwendbar.

34 K. R. SCHERER / H. G. WALLBOTT (Hrsg.): Nonverbale Kommunikation: Forschungsberichte zum Interaktionsverhalten. Weinheim/Basel 1979.

35 R. LAUTMANN: Richterliche Strategien zum Abschluß eines Verfahrens. In: E. BLANKENBURG (Hrsg.): Empirische Rechtssoziologie. München 1975, S. 40–55.

36 K. F. SCHUMANN / G. WINTER: Die Beobachtung im Gerichtssaal. In: E. BLANKENBURG (Hrsg.): Empirische Rechtssoziologie. München 1975, S. 77–95.

Die nichtreaktiven Meßmethoden können also aufgrund der Begrenzungen nicht als ausschließende Alternative zu den reaktiven Verfahren eingesetzt werden, aber sie sind sicherlich, in Verbindung mit den übrigen Instrumenten, eine sehr wichtige und interessante Ergänzung.

### Befragung

### 24.2.3.3.

Trotz ihrer bekannten Schwächen bleibt die Befragung in psychologischen Untersuchungen eines der wichtigsten Instrumente. Vor allem über das subjektive Erleben mit seinen emotionalen und kognitiven Bewertungen läßt sich durch die Befragung Aufschluß gewinnen (*Abb. 11*).

Verfahren 3

*Abb. 11:* Befragung



Eine besondere, standardisierte Art der Befragung sind Tests. Hier wird ein Rorschach-Test durchgeführt, bei dem man auswertet, was der Befragte in einer Reihe von Tintenklecksen „sieht“.

Die empirische Sozialforschung hat eine Reihe von Befragungstechniken entwickelt,<sup>37</sup> die zu unterschiedlichen Zwecken und Zeitpunkten verwandt werden: Es sind dies mehr oder weniger standardisierte bzw. strukturierte *Interviews* und *Fragebogen*.

### Interview

Interviews werden als nichtstandardisierte, halbstandardisierte oder standardisierte Befragungsform eingesetzt.

- Im *nichtstandardisierten Interview* spricht der Interviewer frei mit der Versuchsperson, läßt sich vom Gesprächsfluß leiten, geht auf Dinge ein, die ihm besonders auffallen, usw. Zur Hypothesengewinnung sind solche freien Gespräche äußerst nützlich und notwendig. Für eine gezielte Untersuchung allerdings reichen sie nicht aus, da die Ergebnisse zu sehr vom Untersucher, dessen Interessen, Einstellungen usw. beeinflußt werden können.
- Im *halbstandardisierten Interview* sind vor allem die Themengebiete festgelegt, die in jedem Fall behandelt werden müssen. Je nach Gesprächsablauf können die Abfolge der Themen oder die Art der Fragen und Nachfragen variiert werden.
- Das *standardisierte Interview* legt die Fragen und deren Abfolge verbindlich fest (Tab. 2). Nach Untersuchungen von COX an psychiatrischen Patienten werden damit vollständigere Informationen gewonnen als in der freien Exploration.

Tab. 2: Auszug aus Standardisiertem Interview zur Depression (SID-Kurzfassung) mit Bewertungskategorien für die Antworten

14. Wie empfinden Sie Ihre Leistungsfähigkeit?

sehr gut		gut		mittel		schlecht		sehr schlecht
1	2	3	4	5	6	7	8	9

15. Ermüden Sie leicht?

sehr ausdauernd		ausdauernd		weder noch		leicht ermüdbar		sehr leicht ermüdbar
1	2	3	4	5	6	7	8	9

16. Macht Ihnen jede Kleinigkeit Mühe?

keine		kaum		etwas		viel		sehr viel
1	2	3	4	5	6	7	8	9

Original: H. ELLGRING.

37 Vgl. P. ATTESLANDER: Methoden der empirischen Sozialforschung. Berlin 1969, S. 70–122. – J. van KOOLWIJK / M. WIEKEN-MAYSER (Hrsg.): Techniken der empirischen Sozialforschung. Bd. 1–8. Bd. 4: Erhebungsmethoden: Die Befragung. München 1974. – E. K. SCHEUCH: Das Interview in der Sozialforschung. In: R. KÖNIG (Hrsg.): Handbuch der empirischen Sozialforschung. Bd. 2. Stuttgart 1974, S. 66–190.

Dabei kommen in gleicher Weise Gefühle zur Sprache.<sup>38</sup> Gerade in der Behandlung von gefühlsmäßig besetzten Themen wurde bisher ein Mangel im standardisierten Interview gesehen – der sich in der genannten Untersuchung allerdings nicht finden ließ. Bei der Entwicklung eines Interviews ist besonders auf die Frageform zu achten. So kann man durch die Art der Fragen ein weiches, hartes oder neutrales Interview gestalten. Die Frageform bestimmt auch, wie ausführlich die Antwort der Versuchsperson sein wird. Eine geschlossene Frage („Geht es Ihnen eher gut oder eher schlecht?“) führt zu kürzeren Antworten als eine offene Frage („Wie geht es Ihnen?“). Im Interview bevorzugt man in der Regel offene Fragen.

### Fragebogen

Zu erwähnen sind hier auch zahlreiche Formen von Fragebogen. Zu den Einzelfragen in einem Fragebogen gibt die Person entweder eine zustimmende oder ablehnende Antwort oder stuft die Antwort auf einer Skala ab. Beispiel aus einem Fragebogen zur sozialen Selbstsicherheit: „Ich finde es schwierig, mit einem Fremden ein Gespräch zu beginnen“ (Tab. 3).

Tab. 3: Ausschnitt aus einem Fragebogen zur sozialen Selbst(un)sicherheit

	stimmt gar nicht	stimmt voll- kommen
1. Ich treffe Entscheidungen schnell und sicher	0 . . 1 . . 2 . . 3 . . 4 . . 5	
2. Ich erhalte nicht gern Geschenke	0 . . 1 . . 2 . . 3 . . 4 . . 5	
3. Ich kann Kritik leicht und offen äußern	0 . . 1 . . 2 . . 3 . . 4 . . 5	
4. Ich finde es schwierig, mit einem Fremden ein Gespräch zu beginnen	0 . . 1 . . 2 . . 3 . . 4 . . 5	
5. Ich verlasse mich im allgemeinen auf mein eigenes Urteil	0 . . 1 . . 2 . . 3 . . 4 . . 5	
6. Ich schlucke meinen Ärger immer runter	0 . . 1 . . 2 . . 3 . . 4 . . 5	
7. In Diskussionen fallen mir immer erst nachher die richtigen Argumente ein	0 . . 1 . . 2 . . 3 . . 4 . . 5	
8. Ich habe immer das Gefühl, andere Leute zu belästigen, wenn ich sie um eine Auskunft bitte	0 . . 1 . . 2 . . 3 . . 4 . . 5	
9. Ich unterlasse alles, was Widerspruch herausfordern könnte	0 . . 1 . . 2 . . 3 . . 4 . . 5	
10. Wenn mir jemand ins Wort fällt, fordere ich ihn auf, mich ausreden zu lassen	0 . . 1 . . 2 . . 3 . . 4 . . 5	
11. Ich muß öfters gegen meine Schüchternheit ankämpfen	0 . . 1 . . 2 . . 3 . . 4 . . 5	
12. Es ist mir gleichgültig, was andere Leute über mich denken	0 . . 1 . . 2 . . 3 . . 4 . . 5	

Aus: R. ULLRICH de Muynck / R. ULLRICH: Das Assertiveness-Training-Programm ATP. Einübung von Selbstvertrauen und sozialer Kompetenz. Teil 1. München 1978, Materialmappe: U-Fragebogen (U 1).

Je nachdem, wie differenziert man die Antworten auswerten möchte, können Antwort-Alternativen wie „stimmt – stimmt nicht“ oder eine Abstufung von „stimmt vollkommen“ über „stimmt etwas“ bis zu „stimmt gar nicht“ gewählt werden. Sind Abstufungen bei der Beurteilung vorgenommen, spricht man auch von „Rating-Skalen“.

38 A. COX / D. HOLBROOK / M. RUTTER: Psychiatric interviewing techniques. VI. Experimental study: Eliciting feelings. *British Journal of Psychiatry* 139 (1981), S. 144–152. – A. COX / K. HOPKINSON / M. RUTTER: Psychiatric interviewing techniques. II. Naturalistic study: Eliciting factual information. *British Journal of Psychiatry* 138 (1981), S. 283–291.

Bei einer Befragung, insbesondere zum Erleben von Personen, ist dabei immer zu berücksichtigen, daß die Antwort das Produkt vielfältiger psychologischer Filterprozesse darstellt. Die Antwort wird zum Beispiel durch soziale Normen beeinflusst, durch unvollständige Erinnerungen, durch die Neigung, Dinge zu verallgemeinern, und vieles mehr. Dennoch wird man auf diese Informationsquelle als ökonomisch und vielfältig einsetzbares Instrument nicht verzichten wollen.

#### 24.2.4. Die statistische Auswertung

##### Entscheidung 4

Um Beobachtungsdaten und Meßergebnisse komprimiert darzustellen und zufalls-kritisch zu bewerten, bedient man sich statistischer Methoden. Die Statistik ist nicht auf eine bestimmte Wissenschaft festgelegt, und entsprechend verwenden Ethologen und Psychologen identische Methoden. Einige der Grundlagen seien kurz erläutert. Eine weit verbreitete Meinung lautet, mit Statistik sei „alles zu beweisen“. Falls fehlerhaft angewandt und interpretiert, trifft dies zweifellos zu – aber das ist keine Besonderheit der Statistik.

Individuen handeln in gleichen Situationen unterschiedlich, und ebenso ändert sich das Verhalten des einzelnen Tieres oder einzelnen Menschen im Laufe der Zeit. Aufgrund der inter- und intra-individuellen Variation sind besondere Verfahren nötig, um eine Gruppe oder ein einzelnes Individuum zu charakterisieren. Dies leistet die beschreibende Statistik (*deskriptive Statistik*). Sie ermöglicht, die Variation eines Merkmales in Kennwerte zu fassen, insbesondere einen mittleren Wert und ein Streuungsmaß. Je nach Häufigkeit der einzelnen Varianten eines Merkmales sind manchmal auch weitere Kennwerte nötig, zum Beispiel für die Beschreibung von Asymmetrien einer Verteilung.

Vergleichen wir die Kinderzahl zweier Familien, können wir mit Bestimmtheit sagen, Familie A hat mehr, gleich viel oder weniger Kinder als Familie B. Möchten wir als nächstes zwei Gruppen von Familien vergleichen, beispielsweise Land- und Stadtbewohner, so ist eine derart bestimmte Aussage eigentlich nur möglich, wenn wir sämtliche Land- und Stadtfamilien erfassen, die „Grundgesamtheit“ (*Population*). Psychologen und ebenso Ethologen arbeiten jedoch meist mit „Stichproben“, deren Kennwerte „zufällig“ von denen der Grundgesamtheit abweichen. Ein Vergleich von Stichproben ist nur mit Verfahren der prüfenden Statistik (*Inferenzstatistik*) möglich. Damit kann keine absolut sichere Aussage mehr erreicht werden, sondern es werden Wahrscheinlichkeitsaussagen gemacht.

Die Statistik erlaubt jedoch, das Maß der Unsicherheit einer Aussage zu bestimmen, allerdings nur in bezug auf die Verschiedenheit, nicht aber die Gleichheit von Stichproben. Was heißt das? Jeder statistische Vergleich geht von der „Nullhypothese“ aus, die Stichproben seien gleich. Dem steht die „Alternativhypothese“ der Ungleichheit gegenüber. Es sind nun zwei Fehler möglich: Man kann die Nullhypothese beibehalten, obwohl sie falsch ist („Fehler erster Art“) oder die Nullhypothese verwerfen, obwohl sie richtig ist („Fehler zweiter Art“). Da der Forscher im allgemeinen Unterschiede herausfinden und nicht Gleichheiten feststellen möchte, ist die Statistik auf den Fehler zweiter Art ausgerichtet. Im Resultat-Teil von wissenschaftlichen Berichten findet man häufig Sätze wie: „Der Unterschied ist auf dem 5%-Niveau gesichert.“ (Statt „gesichert“ heißt es oft „signifikant“.) Dies bedeutet, daß der Untersucher die Alternativhypothese eines Unterschiedes zwischen den Stichproben annimmt („der Unterschied ist gesichert“) und die Fehlerwahrscheinlichkeit für diesen Schluß höchstens 5% („auf dem 5%-Niveau“) beträgt.

Beim Milgram-Experiment beispielsweise könnte man inferenz-statistisch prüfen, ob die Durchschnittswerte erteilter Maximalschocks für die vier experimentellen Bedingungen „überzufällig“ verschieden sind, ob also zum Beispiel im Durchschnitt „wirklich“ stärkere Maximalschocks gegeben wurden, wenn sich die Versuchsperson im Fernraum zum „Opfer“ befand statt in Berührungsnähe (Tab. 4).

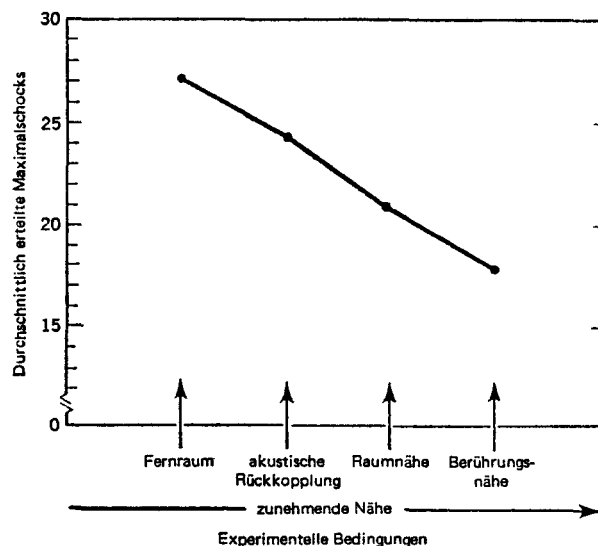


Tab. 4: Zur deskriptiven Statistik des Milgram-Experiments

Schockstufe	Aufschrift und Voltangabe	Experiment 1	Experiment 2	Experiment 3	Experiment 4
		Fernraum (n=40)	Akustische Rückkopplung (n=40)	Raumnähe (n=40)	Berührungsnähe (n=40)
1	Leichter Schock 15				
2	30				
3	45				
4	60				
5	Mäßiger Schock 75				
6	90				
7	105			1	
8	120				
9	Mittlerer Schock 135		1		1
10	150		5	10	16
11	165		1		
12	180		1	2	3
13	Kräftiger Schock 195				
14	210				1
15	225			1	1
16	240				
17	Schwerer Schock 255				1
18	270			1	
19	285		1		1
20	300	5*	1	5	1
21	Sehr schwerer Schock 315	4	3	3	2
22	330	2			
23	345	1	1		1
24	360	1	1		
25	Gefahr! Bedrohlicher Schock 375	1		1	
26	390				
27	405				
28	420				
29	XXX 435				
30	450	26	25	16	12
Durchschnittlich gegebener Maximalschock		27,0	24,53	20,80	17,88
Prozentsatz gehorsamer Versuchspersonen		65,0%	62,5%	40,0%	30,0%

\*Gibt an, daß bei Experiment Nr. 1 fünf Versuchspersonen einen Maximalschock von 300 Volt zufügten.

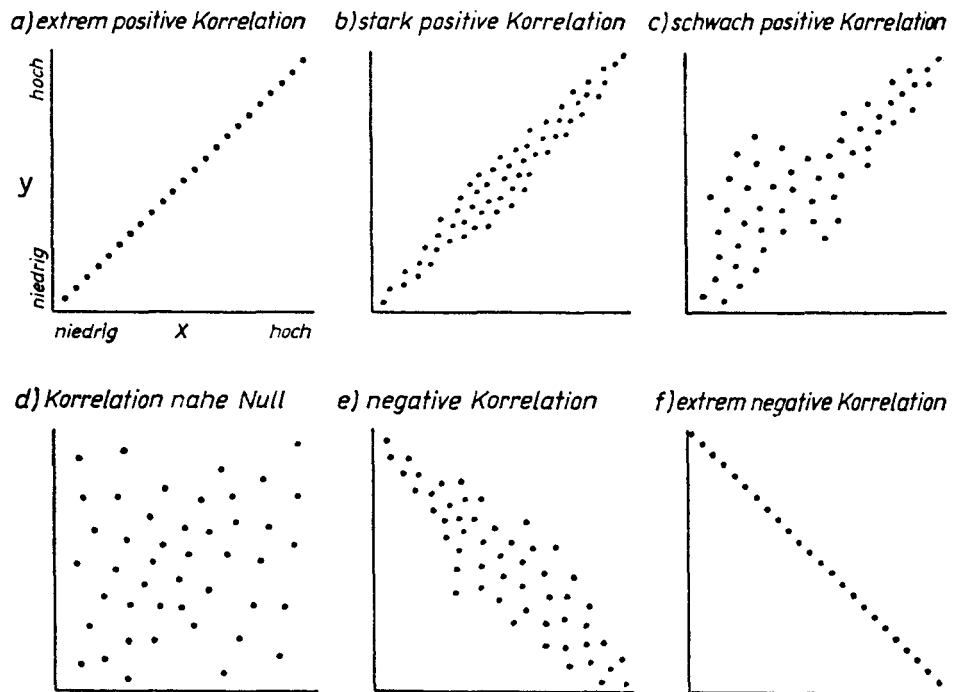
In der Tabelle (oben) sind die Häufigkeiten, das heißt die Anzahl der Versuchspersonen, für die erteilten Maximalschocks in den vier experimentellen Bedingungen angegeben. Ein durchschnittlich gegebener Maximalschock läßt sich für jede Bedingung berechnen und grafisch veranschaulichen (unten).



Aus: S. MILGRAM: Das Milgram-Experiment. Zur Gehorsamsbereitschaft gegenüber Autorität. Reinbek 1974, S. 51f.

Um den Zusammenhang zweier Merkmale zu beschreiben und statistisch zu prüfen, bedient man sich der *Korrelation* ( $r$ ). Eine Korrelation zwischen Physik- und Mathematikschulnoten von  $r=0.85$  beispielsweise besagt, daß beide Merkmale sehr eng miteinander verbunden sind („kovariieren“). Korrelationen können Werte von  $r = -1$  bis  $r = +1$  annehmen, wobei die Korrelation von  $r = 0$  besagt, daß beide Merkmale unabhängig voneinander große oder kleine Werte annehmen können. Eine negative Korrelation besagt, daß große Werte im Merkmal  $x$  mit kleinen Werten im Merkmal  $y$  einhergehen (Abb. 12).

Abb. 12: Korrelationen



Eine Korrelation gibt den Zusammenhang zweier Merkmale in einer Stichprobe wieder. Jeder Punkt im Diagramm kennzeichnet eine Versuchsperson mit ihren Meßwerten auf den beiden Merkmals-Variablen.

- a) Grenzwert  $r = +1$
- b) Punktzahl im Mathematik- und Physiktest
- c) Körpergröße und Gewicht
- d) Intelligenz und Schuhgröße
- e) Vorübung und Fehlerzahl bei Geschicklichkeitsaufgabe
- f) Grenzwert  $r = -1$

Aus: G. CLAUSS / H. EBNER: Grundlagen der Statistik. Zürich/Frankfurt <sup>2</sup>1975, S. 107.

Auch für eine Korrelation ist deren statistische Signifikanz zu prüfen, indem man mit speziellen Prüfverfahren feststellt, ob sie überzufällig von  $r = 0$  verschieden ist. Für Korrelationen wie für Mittelwertvergleiche zwischen Gruppen gilt: Je größer die Stichprobe, desto eher erhält man eine „statistische Signifikanz“. Konkret bedeutet dies, daß bei genügend großer Stichprobe auch minimale Unterschiede zwischen den Gruppenergebnissen statistisch signifikant werden können. Daher geht man in jüngerer Zeit dazu über, nicht nur das Signifikanz-Niveau zu berichten, sondern auch die Größe des Effekts (*effect size*) abzuschätzen, den die gesetzten Bedingungen erzeugt haben. Eine statistische Signifikanz sagt also noch nichts über die inhaltliche Bedeutsamkeit von gefundenen Unterschieden aus. Die Abschätzung der *Effect size* kann hierfür als ein zusätzliches formales Kriterium dienen.

Zusammengefaßt dient die Statistik dazu, Variationen und mittlere Ausprägungen zu beschreiben und Hypothesen anhand von Stichproben zu prüfen. Jede Aussage, die mittels Stichproben gewonnen wird, ist mit einer Unsicherheit behaftet. Für das Akzeptieren eines Unterschiedes zwischen Stichproben kann die Unsicherheit genau berechnet werden, nicht aber für das Akzeptieren der Gleichheit.

Stellen Sie Vor- und Nachteile von Beobachtungs- und Befragungs-Methoden gegenüber.

**Aufgabe 2**

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

### **Das Ergebnis**

**24.3.**

Ist die statistische Auswertung einer Untersuchung abgeschlossen, sind die zu Beginn aufgestellten Hypothesen abschließend im Lichte dieser Daten zu prüfen. Der Untersucher entscheidet, ob die Hypothesen gestützt oder entkräftet werden, und gelangt zu einem veränderten Wissen – das eventuell zu neuen Fragestellungen und Untersuchungen Anlaß gibt.

**Schritt 3**

Ein weiterer, zweiter Aspekt der Bewertung der Ergebnisse einer Untersuchung betrifft deren Gültigkeit und Aussagekräftigkeit – es ist die Frage, ob die Ergebnisse der Studie durch methodische Probleme beeinträchtigt sind.

### **Die Bewertung**

**24.3.1.**

Wie man zu Beginn einer wissenschaftlichen Untersuchung Fragen und Hypothesen formulierte, so sind nun die Ergebnisse im Hinblick auf diese Fragen zu bewerten. Die statistische Prüfung war dabei ein formaler Aspekt der Bewertung. Inhaltlich ist zu fragen, welche neuen Informationen gewonnen wurden und inwieweit sie bisherige theoretische Vorstellungen stützen bzw. mit ihnen nicht vereinbar sind oder auch, inwieweit sie zu neuen theoretischen Überlegungen Anlaß geben.

**Fragenkreis 1**

Nehmen wir das Experiment von MILGRAM, so waren es zwei Ergebnisse, die hervorgehoben wurden: Zum einen die Tatsache, daß unter bestimmten Bedingungen des Autoritätsgefälles (Versuchsleiter vs. naive Versuchspersonen als „Lehrer“) normale, unauffällige Personen Handlungen durchführen, die sie unter anderen Bedingungen für ethisch nicht vertretbar halten würden. Weiterhin zeigte sich als eine ganz wesentliche Variable die soziale Nähe zum „Opfer“: Je näher der „Lehrer“ dem „Schüler“ war, desto eher weigerte er sich, die Bestrafung fortzusetzen bzw. zu verstärken. Aufgrund dieses Ergebnisses lassen sich nun andere Phänomene erklären, die man ansonsten als unverständliche Reaktionen betrachten würde, etwa die Tatsache, daß immer wieder Menschen dazu gebracht werden, andere zu foltern, nur weil sie einen Befehl dazu erhalten.

In jeder der Untersuchungen, die im Rahmen dieses Funkkollegs dargestellt werden, ist die Bewertung der Ergebnisse ein zentraler methodischer Punkt. Strittig

ist dabei häufig, wieweit man die Ergebnisse generalisieren kann und ob Alternativerklärungen möglich sind.

Nehmen wir als Beispiel die Frage, ob mimischer Ausdruck angeboren ist. DARWIN hatte angenommen, daß der Ausdruck kulturinvariant und Teil eines ererbten Repertoires ist.<sup>39</sup> Untersuchungen hierzu betrachteten zum Beispiel das Erkennen des Ausdrucks und fanden in verschiedenen Kulturen ähnliche Zuordnungen von Emotionsbeschreibungen und Gesichtsausdrücken.<sup>40</sup> Eine Alternativerklärung dafür, daß angeborene Prozesse hier eine Rolle spielen, wäre die, daß in den verschiedenen untersuchten Kulturen doch bereits europäisch-amerikanische Einflüsse vorhanden waren. Dies konnte allerdings für einige Kulturen nahezu ausgeschlossen werden. Hinsichtlich der Generalisierung der Aussage scheint es allerdings so zu sein, daß die Kulturinvarianz nur für einige Emotionen gilt, für andere wiederum nicht. In den entsprechenden Theorien spricht man daher von Primär-Emotionen, die beispielsweise Freude, Ärger, Abscheu umfassen, und zusammengesetzten bzw. sozial überformten Emotionen, die kulturspezifische Ausdrucksweisen beinhalten. Die Kulturinvarianz emotionalen Ausdrucks wurde also eingeschränkt auf einen bestimmten Teil der Emotionen.

Der Forscher selbst neigt in der Regel dazu, seine Ergebnisse sehr stark zu generalisieren und Alternativerklärungen für unwahrscheinlich zu halten. Funktion der wissenschaftlichen Diskussion und vor allem auch weiterführender Untersuchungen ist es, die Reichweite von Aussagen immer wieder kritisch zu prüfen.

### 24.3.2. Die Reaktivitätsproblematik

#### Fragenkreis 2

Zunächst muß man davon ausgehen, daß jede empirische Untersuchung grundsätzlich einen Eingriff in das „Untersuchungsfeld“ darstellt. Durch den Meßprozeß kann ein Ergebnis beeinflusst werden, das heißt teilweise künstlich zustande kommen. Man spricht in diesem Fall von einem „Forschungs-Artefakt“. Der Ansatzpunkt methodischer Untersuchungen in der „Artefaktforschung“ ergibt sich aus dem Umstand, daß bei psychologischen Experimenten wie auch bei Befragungen die Erhebung der Daten notgedrungen in einer (sozialen) Interaktionssituation mit all ihren kommunikativen Merkmalen stattfindet. Die Untersuchung wird zum Beispiel in einem Labor von einem Versuchsleiter durchgeführt, der den Versuchspersonen erklärt, was sie im einzelnen machen sollen, oder der Interviewer liest den Befragten einzelne Fragen vor und notiert die Antworten.

Im Gegensatz zur naturwissenschaftlichen Forschung muß man davon ausgehen, daß das zu analysierende „Objekt“ ein wahrnehmendes, interpretierendes und antizipierendes Wesen ist, das wahrscheinlich mehr oder weniger stark auf den Untersuchungsvorgang reagiert: Die Versuchsperson wird unter Umständen versuchen, die Ziele der Untersuchung zu erkennen, und sich dann in der Weise darstellen, die ihr selbst in Anbetracht der vermuteten Ziele erstrebenswert erscheint. Dem Wasser ist es egal, ob die Temperatur gemessen wird oder nicht – um ein Beispiel aus der Physik zu nehmen –, aber Personen, die bewußt wahrnehmen, daß sie als Untersuchungsobjekt an einer (psychologischen) Studie teilnehmen, werden vielleicht den Meßvorgang als solchen stark beeinflussen – man spricht in diesem Zusammenhang von der „Reaktivitätsproblematik“. Auch bei Untersuchungen an Tieren sind solche Einflüsse bekannt.

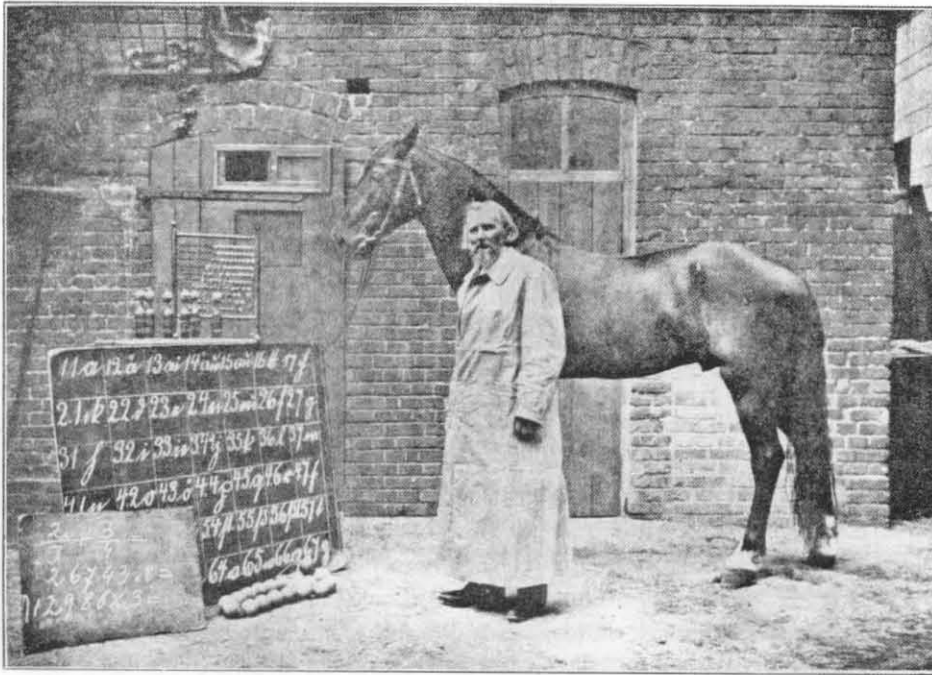
Der „Kluge Hans“, ein Pferd des Herrn VON OSTEN, ist das erste berühmte wissenschaftlich untersuchte Beispiel. Das Pferd konnte „rechnen“, indem es Ergebnisse mit dem Fuße stampfend mitteilte (*Abb. 13*). Der Psychologe PFUNGST fand 1924 heraus, daß das Pferd nicht etwa rechnete, sondern genau die Reaktionen des Publikums beobachtete.<sup>41</sup> Offensichtlich schauten die Zuschauer zunächst gebannt auf den zählenden Huf und lehnten sich vielleicht unbewußt zurück, wenn das richtige Ergebnis erreicht war.

39 Ch. DARWIN: *The Expression of the Emotions in Man and Animals*. London 1872.

40 P. EKMAN: *Cross cultural studies of facial expression*. In: P. EKMAN (Hrsg.): *Darwin and Facial Expression*. New York 1973, S. 169–222.

41 O. PFUNGST: *Das Pferd des Herrn v. Osten*. Leipzig 1907. (2. Auflage unter dem Titel „Der kluge Hans“. Frankfurt 1977).

Abb. 13: „Der kluge Hans“



Aus: O. PFUNGST: Der kluge Hans. Ein Beitrag zur nicht-verbalen Kommunikation. Frankfurt 1977, S. 36, Vorwort.

Wenn man dem Pferd die Augen verband, wie es PFUNGST mit Zustimmung des Pferdebesitzers tat, so war es mit den Rechenkunststücken vorbei. Dieses Experiment wurde später von ROSENTHAL aufgegriffen, der zeigte, daß auch Ratten beim Labyrinthlernen durch das unbewußte Verhalten des Versuchsleiters positiv oder negativ beeinflusst werden.<sup>42</sup>

ORNE, einer der bekanntesten „Artefaktologen“, entwickelte in diesem Zusammenhang das Konzept der „Demand characteristics“.<sup>43</sup> Er ging davon aus, daß die Versuchspersonen im Labor die jeweilige Untersuchung als eine Problemlösungssituation interpretieren: Sie versuchen, die Hypothese des Versuchsleiters zu erraten. Sie werden deshalb alle Hinweiszeichen als Anforderungscharakteristika der Situation (*demand characteristics*) aufmerksam registrieren, die ihnen dabei behilflich sein können, die an sie gestellten Erwartungen zu erkennen.

Eine sehr wichtige Rolle spielt dabei der Versuchsleiter als wohl bedeutsamste Quelle für mögliche Demand characteristics. Es ist zweifelsohne das Verdienst von ROSENTHAL gewesen, als erster systematisch derartige Versuchsleitereffekte näher analysiert zu haben.<sup>44</sup> Bekannt wurde insbesondere der „Versuchsleiter-Erwartungs-Effekt“ („Rosenthal-Effekt“), wonach Versuchsleiter mehr oder weniger unbewußt mit Hilfe verbaler und non-verbaler Signale den Versuchspersonen die eigene Hypothese mitteilen. Die Erwartungen werden damit zu „sich selbst erfüllenden Prophezeiungen“ (*self-fulfilling prophecies*).

Die Versuchsleiter-Erwartungseffekte, aber auch die sonstigen Demand-characteristics, dürfen allerdings in ihrer tatsächlichen Wirkung nicht überschätzt werden:

42 R. ROSENTHAL: On the social psychology of the psychological experiment: The experimenters hypotheses as unintended determinant of experimental results. *American Scientist* 51 (1963), S. 268–283.

43 M. T. ORNE: On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist* 17 (1962), S. 776–683.

44 R. ROSENTHAL: *Experimenter Effects in Behavioral Research*. New York 1966 (2., erweiterte Auflage 1976).

Wie die Versuchspersonen auf die erschlossenen Inhalte der Untersuchung reagieren, wird wesentlich bestimmt von den Motiven und Selbstdarstellungsabsichten, die sie selbst in die Untersuchungssituation einbringen. Die meisten wollen dabei, den bisherigen Befunden zufolge, in erster Linie positiv bewertet werden. Diese Personen gehen mit der Erwartung ins Labor, daß trotz aller Täuschungsversuche letztlich ihre Fähigkeiten und Persönlichkeitseigenschaften (wie z. B. Intelligenz, psychische Gesundheit) bewertet werden sollen. Sie bemühen sich deshalb, durch optimale Selbstdarstellungsstrategien möglichst gut abzuschneiden.

Reaktivitätseffekte entstehen nicht nur im Labor, sondern ebenso bei Befragungen. Die Befragten wissen, daß das Gespräch Teil einer wissenschaftlichen Studie ist, und werden sich entsprechend auf diese Situation einstellen. Wer führt dieses Interview mit welcher Absicht durch? Welche Konsequenzen haben die Ergebnisse für mich persönlich? Welche Antworten erwartet der Interviewer?

Neben derartigen Spekulationen um die Zielsetzung der Befragung werden die Befragten häufig auch Selbstdarstellungsstrategien praktizieren, um sich als „normaler“ Bürger auszugeben. Je nachdem, wie stark bestimmte Themenstellungen tabuisiert oder emotional verankert sind, werden die Antworten entsprechend verzerrt sein.

Werden Personen etwa befragt, wie oft sie in der Woche baden, dann werden sie sich bei ihren Antworten teilweise danach richten, was sie glauben, wie oft *man* im Laufe einer Woche baden sollte. Es wäre naiv zu glauben, daß die Antworten etwas mit der tatsächlichen Badehäufigkeit zu tun haben. Es gäbe im übrigen seit vielen Jahren akute Gebrauchswasser- not, wenn die bislang ermittelten Durchschnittswerte zu dieser Frage stimmten. Oder fragt man Eltern beispielsweise nach dem durchschnittlichen Fernsehkonsum ihrer Kinder, dann werden in diesem Fall eher ideale Normen anstatt tatsächliche Fernsehzeiten erfaßt.

Reaktivitätseffekte sind schließlich auch bei Feldexperimenten in psychologischen Anwendungsgebieten nachgewiesen worden.

Bei einer arbeitspsychologischen Untersuchung über den Zusammenhang zwischen Lichtstärke am Arbeitsplatz und Leistungsfähigkeit des einzelnen Mitarbeiters hatte man in den 30er Jahren – zur Überraschung der Forscher – festgestellt, daß die Leistungsfähigkeit sowohl bei Erhöhung als auch bei Reduzierung der Lichtstärke zunahm.<sup>45</sup> Die Ergebnisse erwiesen sich insofern als Artefakte, als die Leistungssteigerung nicht als Folge der Variation der Arbeitsplatzbedingungen (der Lichtstärke) erfolgte, sondern vielmehr darauf zurückzuführen war, daß die Personen auf die Studie als solche reagiert hatten. Sie wußten, daß ihr Verhalten von Wissenschaftlern registriert wurde, was bis dahin nicht üblich gewesen war, und allein dieses „erhabene Gefühl“, zu dem „ausgewählten“ Personenkreis zu gehören, bewirkte die Leistungssteigerungen, was auch immer die Forscher in ihrer Umgebung veränderten.

Psychologen interessieren sich im menschlichen Bereich unter anderem für das, was sich im Bewußtsein der Menschen in bestimmten Situationen abspielt, und gerade dieses Bewußtsein stellt sich also, so paradox es klingen mag, als eines der größten Probleme bei empirischen Untersuchungen heraus: Wann immer eine Person weiß, daß sie an einer psychologischen Untersuchung teilnimmt, wird sie auf den Meßvorgang als solchen reagieren, und die Untersuchung wird mit einem Reaktivitätseffekt belastet. Aber auch im Tierexperiment ist mit solchen Effekten zu rechnen, und selbst die bloße Tatsache der Beobachtung kann das Verhalten des Beobachtungsobjektes unbeabsichtigt beeinflussen. Wenn auch keine vollständige Lösung dieses Problems verfügbar ist, so sollte sich der Forscher in jedem Fall über diese Einflußgrößen bewußt sein.

45 F. ROETHLISBERGER / W. DICKSON: Management and the Worker. Cambridge, Mass. 1939.

### Möglichkeiten und Grenzen einer systematischen Artefaktkontrolle

### 24.3.3.

#### Fragenkreis 3

Angesichts der geschilderten Reaktivitätsprobleme ist es verständlich, daß bei psychologischen Untersuchungen bzw. bei Befragungen die Personen oft über die wahre Zielsetzung im unklaren gelassen, ja durch entsprechende Ablenkungsmanöver getäuscht werden. Dieser Umstand dürfte für das Image derartiger Studien nicht gerade förderlich gewesen sein, und die meisten Versuchspersonen gehen (aufgrund der Kenntnisse psychologischer Experimente aus den Massenmedien) mit der Erwartung in ein psychologisches Labor, daß die Täuschung sozusagen zum „Geschäft“ gehört.<sup>46</sup> Die Neugier der Versuchspersonen wird dadurch aber nur noch stärker entfacht, und die Reaktivitätsproblematik ist damit beileibe nicht aufgehoben. Die naheliegende Frage an die Artefaktforschung lautet also, mit welchen Gegenstrategien diese „unerwünschten“ Prozesse kontrolliert werden können. Die Versuche zur Beseitigung dieser „Störfaktoren“ setzen konsequenterweise dort an, wo die Fehlerquellen lokalisiert wurden, also bei der Person des Versuchsleiters und bei den Versuchspersonen.

Was die Maßnahmen bezüglich des Versuchsleiters betrifft, so wurden sehr unterschiedliche Erfahrungen gemacht. Der Einsatz möglichst vieler Versuchsleiter zur Erreichung eines gewissen Fehlerausgleichs und die Verwendung von Erwartungskontrollgruppen (durch unterschiedliche Manipulation der Erwartungen verschiedener Versuchsleiter-Stichproben) haben sich nicht als wirksame Strategien durchsetzen können. Die Durchführung von „Doppelblindversuchen“ hingegen, bei denen die Versuchsleiter nicht über jeweilige experimentelle Bedingungen informiert werden, ist zu einem Standardverfahren bei jeder Art von Medikamentenprüfung geworden. Am effizientesten scheinen ein intensives Training der Versuchsleiter und eine begleitende Kontrolle ihres Verhaltens während der Versuche zu sein. Problematisch ist dabei allerdings, daß die in Frage kommenden kritischen Versuchsleiter-Aktivitäten nur teilweise bekannt, oft sehr subtil und damit kaum registrierbar und meist erst im nachhinein aus dem Gesamtzusammenhang des Kommunikationsprozesses heraus interpretierbar sind. In einigen Studien erzielten im übrigen gerade die zuvor trainierten Versuchsleiter besonders starke, zum Teil gegenläufige Erwartungseffekte, indem sie ihren Versuchspersonen gleichsam „Anti-Hypothesen“ nahelegten. Die Möglichkeiten einer effizienten Artefaktkontrolle beim Versuchsleiter sind also begrenzt.

Bleibt bestenfalls die Möglichkeit, anhand nach-experimenteller Befragungen oder aufgrund von Simulationsstudien (wie sie von ORNE vorgeschlagen wurden) die wichtigsten Demand characteristics zu identifizieren.<sup>47</sup> Aber selbst wenn man die Laborsituation so gestaltet, daß die hypothesenfördernden Hinweiszeichen auf ein Minimum reduziert werden, muß man davon ausgehen, daß die Gesamtsituation von jeder Versuchsperson in eigener Weise strukturiert und interpretiert wird. Mit anderen Worten: Unkontrollierbare verhaltensrelevante Hypothesen können nicht verhindert werden.

Ob die Versuchsperson die richtige oder aber eine völlig falsche Hypothese aufstellt, ist im Grunde genommen nebensächlich – entscheidend ist, daß sie sich, von dieser Hypothese ausgehend, in einer besonderen Art und Weise darzustellen versucht.

Bei einem Intelligenztest mag dies kein Problem darstellen, denn die Versuchsperson weiß, daß ihre Intelligenz gemessen werden soll. In einem sozialpsychologischen Experiment zur Untersuchung prosozialen Verhaltens (Hilfeleistung) hingegen kann die Versuchsperson unter Umständen einer anderen Person behilflich sein, weil die Demonstration eines solchen

46 S. M. JOURARD: Brief einer Vp an einen V1. *Gruppendynamik* 4 (1973), S. 27–30.

47 M. T. ORNE: Demand characteristics and the concept of quasi-controls. In: R. ROSENTHAL / R. L. ROSNOW (Hrsg.): *Artifact in Behavioral Research*. New York 1969, S. 143–179.

sozial erwünschten Verhaltens offensichtlich zu einer optimalen Selbstdarstellung gehört. Ob die gleiche Person auch sonst (in einer „reaktivitätsfreien“ Situation) generell hilfreich ist, läßt sich aus derartigen Experimenten wohl kaum ablesen.

Bei Befragungen wurden zur Kontrolle der Verzerrungen durch die Versuchspersonen „Lügenskalen“ entwickelt, die zusätzlich zu den übrigen Fragen in den Fragebogen eingebaut werden. Mit solchen Skalen soll die Tendenz, sozial erwünschte Antworten zu geben, erfaßt werden, und je nach Punktzahl werden einzelne Personen entweder aus der Stichprobe herausgenommen, oder aber die Antworten auf die übrigen Fragen werden entsprechend gewichtet bzw. relativiert. Der Grundgedanke bei der Entwicklung solcher Skalen ist einfach: Die Versuchsperson wird beispielsweise gefragt, ob sie irgendwann im Leben einmal etwas gestohlen hat, und man geht davon aus, daß eigentlich 99% der Bevölkerung diese Frage bejahen müßten (und sei es nur, daß man in der Schule einem Kameraden einen Radiergummi „geklaut“ hat). Wenn nun der Befragte diese Frage verneint, dann versucht er sich offensichtlich positiv darzustellen. Wiederholt sich diese Strategie bei allen 20 Fragen, dann erzielt er die maximal erreichbaren 20 Punkte auf der Lügenskala, und der Verdacht liegt nahe, daß diese Verhaltenstendenz auch bei den übrigen Fragen wirksam gewesen ist.

Diese Kontrolltechnik anhand derartiger Lügenskalen ist schon seit vielen Jahrzehnten bekannt.<sup>48</sup> Ihre Brauchbarkeit ist allerdings umstritten. Als „Spezialist“ für Reaktivitätseffekte, die Sie nach der Lektüre der vorangegangenen Kapitel geworden sind, werden Sie bestimmt auch denken, daß gerade solche Lügenskalen einen hohen Aufforderungscharakter haben. Der Kandidat wird sich angesichts der seltsamen Fragestellungen Gedanken über die wahre Absicht dieser Sätze machen und entsprechende Vermutungen entwickeln. Die wahre Tendenz, sich positiv darzustellen, äußert sich unter diesen Umständen vielleicht gerade in einem niedrigen Punktwert. Die Meßergebnisse auf der Lügenskala sind damit aber nicht mehr einheitlich interpretierbar, ihr praktischer Wert ist stark eingeschränkt.

Die Reaktivitätseffekte bei psychologischen Untersuchungen lassen sich also insgesamt gesehen nur teilweise kontrollieren; sie gehören im Grunde genommen als wesentlicher Bestandteil zur Studie dazu. In diesem Sinne handelt es sich also nicht um Störgrößen, die ausgeschlossen, sondern um wichtige Aspekte, die beachtet und ebenfalls erfaßt werden müssen. Bei psychologischen Labor-Experimenten bedeutet dies, daß man zusätzlich zu dem interessierenden Verhalten zumindest im nachhinein die Interpretation der Betroffenen analysieren sollte, um ihr Verhalten auf dem Hintergrund der Hypothesen verstehen zu können. Bei Befragungen muß man grundsätzlich davon ausgehen, daß die Befragten das sagen, wovon sie glauben, daß es den allgemeinen gesellschaftlichen Vorstellungen oder den Normen einer spezifischen Gruppe entspricht. Insofern handelt es sich um ein Erhebungsinstrument zur Erfassung von Werthaltungen, die in einer spezifischen Situation (nämlich der Interaktion zwischen Versuchsperson X und Interviewer Y) geäußert werden. Zusammenfassend betrachtet, läßt sich also aus der Reaktivitätsproblematik die Konsequenz ableiten, daß die Befunde von psychologischen Untersuchungen aufgrund solcher Effekte nicht etwa unbrauchbar sind, sondern im Sinne dieses Phänomens entsprechend interpretiert werden müssen.

### Aufgabe 3

Welche Möglichkeiten der Artefakt-Kontrolle sehen Sie bei einer Beobachtungs-Untersuchung von kindlichem Verhalten?

.....

.....

.....

<sup>48</sup> H. J. HENNING / K. MUTHIG: Grundlagen konstruktiver Versuchsplanung. München 1979.



**Ethische Fragen der Forschung****24.4.****Konflikt**

Experimente mit Tieren sind in letzter Zeit zunehmend in Verruch geraten, wobei vor allem die Versuchsreihen bei der Überprüfung von neuen Medikamenten im Mittelpunkt der Kritik stehen. „Militante“ Tierschützer haben kürzlich in der Freien Universität Berlin zahlreiche „Versuchskaninchen“ in die Freiheit gelassen, um sie so nicht länger den wissenschaftlichen Torturen auszusetzen. Grundsätzlich berührt dieser Aspekt alle psychologischen und ethologischen Tierexperimente, bei denen die Versuchstiere notgedrungen zum Zwecke der Beobachtung in einen Käfig eingesperrt werden. Die ethische Problematik stellt sich erst recht, wenn den Tieren operativ ein Sendegerät eingebaut oder Teile des Gehirns herausgenommen werden, um entsprechende wissenschaftliche Studien durchführen zu können.

Ethische Probleme ergeben sich aber selbstverständlich auch im Humanbereich. So sind zum Beispiel in den letzten Jahren die Persönlichkeitstests im Rahmen der Personalauslese heftig kritisiert worden, da sie den verfassungsmäßig geregelten Schutz der Persönlichkeit verletzen können.<sup>49</sup> Aber auch die in der herkömmlichen Laborforschung praktizierte Täuschung der Versuchspersonen zur angeblichen Vermeidung von Reaktivitäts-Effekten ist unter ethischen Gesichtspunkten fragwürdig. Wird das Vertrauen der naiven Versuchspersonen nicht über Gebühr mißbraucht, wenn die Zielsetzung eines Experimentes verschwiegen wird? Darf das Verhalten der Personen ohne deren Wissen mit Video aufgezeichnet und später analysiert werden? Laborexperimente führen gelegentlich auch zu extremen Stresssituationen und zu traumatischen Erlebnissen. Woher nimmt der Forscher das Recht, Menschen derartigen Situationen auszusetzen?

Im „Elektro-Schock“-Experiment von MILGRAM mußten die Versuchspersonen eine andere, angebliche Versuchsperson (tatsächlich ein vorinstruierter Helfer des Versuchsleiters) mit Elektroschocks für Fehler in einem Lernprozeß bestrafen. Viele Versuchspersonen „verabreichten“ den Lernenden die maximale Stärke von 450 Volt, und die Konfrontation mit ihrem eigenen zum Teil „mörderischen“ Verhalten hat zwangsläufig zu einer erheblichen psychischen Belastung geführt, die noch einige Zeit nach Abschluß des Experimentes nachgewirkt haben dürfte:

„Ich glaube, er will sich in Verbindung setzen, er klopft ... Also, es ist nicht fair, dem Mann Schocks zu geben ... Das sind scheußlich hohe Spannungen. Ich glaube, die Sache ist nicht sehr human ... O nein, ich kann da nicht weitermachen; nein, das ist nicht recht. Es ist ein verdammt blödes Experiment. Der Mann dort drüben hat doch Schmerzen. Nein, ich will nicht mehr weitermachen. Das Ganze ist verrückt.“ (Die Versuchsperson weigert sich, weitere Schocks zuzufügen.)

„Der klopft doch da drüben. Ich mach' nicht mehr mit. Ich möcht' ja schon weitermachen, aber ich kann das einem Menschen nicht antun ... Tut mir leid, aber ich kann das einem Menschen nicht antun. Ich bring ihm noch einen Herzschaden bei. Nehmen Sie Ihr Geld wieder zurück ... Nein, wirklich, ich könnte das nicht tun.“<sup>50</sup>

Das Experiment von MILGRAM hat angesichts der weitreichenden Publizität innerhalb der Psychologie eine heftige Kontroverse zur ethischen Verantwortung des Forschers ausgelöst. Er hat im Anhang seines Buches zur Rechtfertigung seines Experiments selbst Stellung bezogen:

„Ein wichtiger Aspekt des Verfahrens ergab sich nach Beendigung der experimentellen Sitzung. Alle Versuchspersonen wurden nach dem Experiment sorgfältig behandelt. Wie sie im einzelnen aufgeklärt wurden, variierte von Versuch zu Versuch und entsprechend unserer zunehmenden Erfahrung. Zumindest wurde allen erklärt, daß das Opfer nicht wirklich gefährliche Elektroschocks erhalten hatte. Jede Versuchsperson hatte Gelegenheit zu einer freundschaftlichen ‚Aussöhnung‘ mit dem unbeschädigten Opfer und zu einer ausgedehnten Diskussion mit dem Versuchsleiter. Gehorsamen Versuchspersonen wurde versichert, daß ihr Verhalten vollkommen normal gewesen sei und daß andere Teilnehmer die gleichen

49 Vgl. S. VON PACZENSKY: Der Testknacker. Wie man Karriere-Tests erfolgreich besteht. Reinbek 1976.

50 S. MILGRAM: Das Milgram-Experiment. Zur Gehorsamsbereitschaft gegenüber Autorität. Reinbek 1974, S. 48.

Konflikt- und Spannungsgefühle wie sie durchgemacht hätten. Man erklärte ihnen, sie würden alle nach Abschluß der Versuchsreihe einen umfassenden Bericht erhalten. In manchen Fällen fanden mit einzelnen Versuchspersonen zusätzlich eingehende und langdauernde Diskussionen über das Experiment statt.

Nachdem die Versuchsreihe komplett war, erhielten die Versuchspersonen einen schriftlichen Bericht mit den Einzelheiten des experimentellen Verfahrens und den Resultaten. Auch hier waren ihre Rolle im Experiment und ihr Verhalten dabei auf eine rücksichtsvolle Weise dargestellt. Alle Versuchspersonen erhielten einen Anschlußfragebogen über ihre Teilnahme an dem Forschungsunternehmen, der ihnen erneut ermöglichte, ihre Gedanken und Gefühle zu ihrem Verhalten zu äußern.

Die Antworten auf den Fragebogen bestätigten meinen Eindruck, daß die Teilnehmer das Experiment positiv beurteilten.

Tabelle 8: Auszug aus dem Fragebogen, der im Anschluß an die Gehorsamsexperimente vorgelegt wurde

Nachdem ich den Bericht gelesen und alles erwogen habe ...	Ungehorsame	Gehorsame	Alle
1. bin ich sehr froh, an dem Experiment teilgenommen zu haben	40.0%	47.8%	43.5%
2. bin ich froh, an dem Experiment teilgenommen zu haben	43.8%	35.7%	40.2%
3. tut es mir sehr leid, an dem Experiment teilgenommen zu haben	15.3%	14.8%	15.1%
4. tut es mir leid, an dem Experiment teilgenommen zu haben	0.8%	0.7%	0.8%
5. bin ich weder froh, noch tut es mir leid, an dem Experiment teilgenommen zu haben	0.0%	1.0%	0.5%

Ein Jahr nach der Teilnahme schrieb eine Versuchsperson: ‚Dieses Experiment hat die Überzeugung in mir bekräftigt, daß der Mensch es vermeiden sollte, seinem Mitmenschen Schaden zuzufügen, selbst auf das Risiko hin, dadurch Autorität zu verletzen.‘

Eine andere gab an: ‚Für mich unterstrich das Experiment ... das Ausmaß, in dem jeder Einzelne eine feste Grundlage haben oder finden sollte, auf der seine Entscheidungen beruhen sollten, gleichgültig, wie unwichtig sie zu sein scheinen. Ich glaube, die Menschen sollten tiefer über sich selbst und über ihre Beziehung zu ihrer Umwelt und zu anderen Menschen nachdenken. Wenn dieses Experiment dazu dient, Menschen aus ihrer behaglichen Selbstgefälligkeit wachzurütteln, dann hat es seinen Zweck erfüllt.‘ [...]

Die wesentliche moralische Rechtfertigung für ein Verfahren, wie es in meinem Experiment Verwendung fand, besteht darin, daß es von den Teilnehmern für gerechtfertigt gehalten wird. Überdies war es gerade die Eindringlichkeit dieser Erfahrung, die uns die wesentliche moralische Berechtigung zur Fortsetzung der Experimente lieferte.“<sup>51</sup>

Die Täuschung der Versuchspersonen läßt sich natürlich mit dem Hinweis rechtfertigen, daß nur so diese dramatischen Konformitätskonflikte nachweisbar sind und daß die psychische Betroffenheit geradezu erwünscht sei, um einen entsprechenden Reflexionsprozeß in Gang zu setzen. Die Gegner solcher Experimente halten derartige Manipulationen für eine „Antitherapie“, die in krassem Gegensatz zur sonstigen therapeutischen Situation steht.<sup>52</sup> Sie lehnen jegliche Form der Täuschung kategorisch ab.

Unter diesem Gesichtspunkt sind in den letzten Jahren die ethischen und rechtlichen Probleme auch der nichtreaktiven Messungen in den Mittelpunkt der Aufmerksamkeit gerückt.<sup>53</sup> Man muß sich in der Tat fragen, wieweit Personen außerhalb des

51 S. MILGRAM: Das Milgram-Experiment (s. *Anm.* 50). Anhang I: Ethische Probleme in der wissenschaftlichen Forschung, S. 221–231.

52 J. SEEMAN: Deception in psychological research. *American Psychologist* 24 (1969), S. 1025–1028.

53 I. SILVERMAN: Nonreactive methods and the law. *American Psychologist* 30 (1975), S. 764–769.

Labors ohne ihr Wissen beobachtet bzw. fotografiert oder gefilmt werden dürfen. Ist zulässig, daß Sozialforscher, unter Betten versteckt, die Gespräche der Studenten belauschen, Telefongespräche abhören, Notsituationen simulieren, eine Puppe, die als Attrappe für einen Suizidanten dient, an einer Autobahnbrücke aufhängen oder das Verhalten in öffentlichen Bedürfnisanstalten analysieren?

Leider werden die Eingriffe in das Persönlichkeitsrecht, die möglichen physischen wie die nicht so offenkundigen, aber unter Umständen um so schwerwiegenderen psychischen Schädigungen als Folge derartiger Manipulationen, allzuoft mißachtet. Die Problematik wird leichtfertig im Zuge der Proklamierung eines falschverstandenen Wissenschaftsethos unter den Tisch gekehrt. Angesichts solcher Praktiken darf man sich natürlich nicht wundern, wenn dann in der Öffentlichkeit die Tätigkeit psychologischer Forscher mit „Fallen-Stellen“ von „Voyeuren“ und „Partisanentätigkeit“ gleichgesetzt wird. Tatsächliche Notsituationen werden vielleicht in Folge einer solchen Einstellung sogar irrtümlicherweise mit einem „harmlosen“ Feldexperiment verwechselt.

Die ethischen Fragen werden in der Praxis bisher in eigener Regie des verantwortlichen Forschers gelöst, wobei er sich aus seiner Sicht (extrem ausgedrückt) für „unmoralische“, aber wissenschaftlich brauchbare und gültige Ergebnisse oder ethisch einwandfreie, aber nutzlose Befunde entscheiden muß. Letztere werden allgemein als der wissenschaftlichen Karriere kaum förderlich eingestuft. Man muß sich also vorerst mit der Warnung begnügen, daß man insbesondere bei der Verwendung nichtreaktiver Meßverfahren ständig Gefahr läuft, eingesperrt zu werden.<sup>54</sup> Je nach Untersuchung kann es auch passieren, daß man, wie im Falle der interessanten Studie von VALENTINI und DI MEGLIO, von der katholischen Kirche exkommuniziert wird.<sup>55</sup>

Fassen wir abschließend noch einmal zusammen: Jede wissenschaftliche Untersuchung durchläuft verschiedene Phasen, in denen geplant wird, Beobachtungen und Experimente durchgeführt und schließlich die Ergebnisse ausgewertet und interpretiert werden. Dieser logische Ablauf wird zwar in der Realität keineswegs immer strikt eingehalten. So würde man idealerweise zunächst eine Hypothese aus einer Theorie ableiten, dann für die Überprüfung der Hypothese ein Experiment entwickeln, usw. Tatsächlich ist der Ablauf ähnlich den Phasen kindlicher Entwicklung. Sie treffen im Prinzip zu, im konkreten Einzelfall sieht es hingegen meist anders aus. Da stellt man bereits nach wenigen Beobachtungen fest, daß die ursprüngliche Hypothese nicht sehr glücklich formuliert war, formuliert daher neu, usw. Dennoch sollte aus dem dargestellten Ablauf der Weg wissenschaftlicher Verhaltensuntersuchung hervorgehen. Es wurden zudem einige der methodischen Probleme aufgezeigt, die eine systematische Erfassung des Verhaltens mit sich bringt. Schließlich müssen auch immer ethische Fragen berücksichtigt werden, die sich aus den wissenschaftlichen Eingriffen in das Verhalten von Mensch und Tier ergeben.

54 S. BOCHNER: The unobtrusive method in experimental social psychology. Unveröffentlichtes Manuskript 1972.

55 R. VALENTINI / C. DI MEGLIO: Zit. nach N. VORDERMANN: Sie werden eine Tochter des Teufels sein. *Der Stern* 14/1973, S. 202–203.