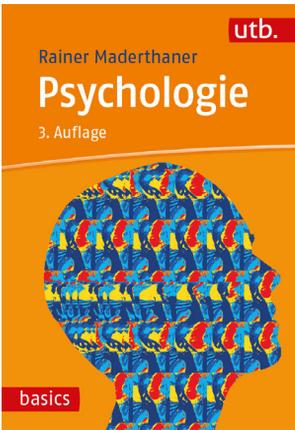


Lernunterlage 1 für Prüfungsteil A zum Aufnahmeverfahren

Psychologie

Studienjahr 23/24

Die Kapitel zur Vorbereitung auf die Aufnahmeprüfung für das Studium der Psychologie werden Ihnen durch eine Initiative des Verfassers, der Universitäten Wien, Salzburg, Graz und Innsbruck sowie des facultas Universitätsverlages zur Verfügung gestellt und sind aus folgendem Buch entnommen:



Rainer Maderthaner

Psychologie

3., vollständig überarbeitete und ergänzte Auflage

utb:facultas 2021, 482 Seiten

ISBN: 978-3-8252-5540-4

Definition, Ziele und Positionen der Psychologie

2

Inhalt

- 2.1 Definitionen von Psychologie

- 2.2 Allgemeine Zielsetzungen wissenschaftlicher Psychologie
 - Beschreiben
 - Erklären
 - Vorhersagen
 - Verändern

- 2.3 Kontroverse Grundannahmen der Psychologie
 - Leib – Seele
 - Anlage – Umwelt
 - Vergangenheit – Gegenwart
 - Freier Wille – Determiniertheit
 - Bewusst – unbewusst
 - Allgemeingültigkeit – Einzigartigkeit
 - Wertfreiheit – Wertbekenntnis
 - Objektivität – Subjektivität
 - Zergliederung – Ganzheitlichkeit
 - Statik – Dynamik
 - Quantitativ – qualitativ

- 2.4 Gegenwärtige Forschungsorientierungen der Psychologie

Definitionen von Psychologie

2.1

Das Wort **Psychologie** bedeutet, wie erwähnt, „Seelenkunde“ oder „Seelenlehre“ (griech. „psyche“: Hauch, Leben, Seele; griech. „logos“: Wort, Begriff). Die Auffassungen darüber, was unter Seele verstanden wird, unterscheiden sich jedoch ziemlich. Nachfolgend

sollen einige innerhalb des Wissenschaftsfaches Psychologie verbreitete Definitionen und Umschreibungen für „Psychologie“ präsentiert werden.

„Die Psychologie ist eine empirische Wissenschaft. [...] Ihr Gegenstand ist das (zumeist menschliche) Erleben und Verhalten, ihr Ziel ist es, allgemeingültige Aussagen über diesen Gegenstand zu machen – ihn zu beschreiben, beobachtbare Regelmäßigkeiten und Zusammenhänge aufzudecken, diese zu erklären, und womöglich Vorhersagen zu machen“ (Hofstätter & Wendt, 1974, 1). In ähnlicher Weise versteht Traxel (1974, 15) die Psychologie als Erfahrungswissenschaft, die als ein „System methodisch gewonnener Aussagen über einen bestimmten Gegenstand“ zu definieren ist.

Merksatz

Psychologie untersucht die Zustände und Veränderungen des Verhaltens, des Erlebens und des Bewusstseins.

Als zentral für die Definition von Psychologie wird oft die Angabe des Forschungsgegenstands angesehen, mit dem sich das Fach zu beschäftigen hat. Bourne und Ekstrand (1992, 2) formulieren: „Die Psychologie ist die wissenschaftliche Er-

forschung von Verhalten.“ Bei dieser breiten Definition könnte das Missverständnis entstehen, es sei nur das „äußere“ (beobachtbare) Verhalten gemeint. In Rohrachers international viel beachtetem Werk „Einführung in die Psychologie“ gelten dagegen die bewussten Prozesse mit ihren Auslösern und Effekten als Hauptcharakteristikum des Forschungsfelds der Psychologie: „Psychologie ist die Wissenschaft, welche die bewußten Vorgänge und Zustände sowie ihre Ursachen und Wirkungen untersucht“ (Rohracher, 1965, 7). Hier werden die zahlreichen unbewussten, automatisch ablaufenden psychischen Vorgänge noch vernachlässigt, zumindest aber ergibt sich eine Abgrenzung zu anderen Humanwissenschaften.

Zimbardo und Gerrig (1999, 2) definieren: „Gegenstand der Psychologie sind Verhalten, Erleben und Bewusstsein des Menschen, deren Entwicklung über die Lebensspanne und deren innere (im Individuum angesiedelte) und äußere (in der Umwelt lokalisierte) Bedingungen und Ursachen.“ Diese Definition ist bereits spezifischer. Die Bedeutung „innerer“ (introspektiver) Prozesse für die psychologische Forschung – der europäischen Tradition entsprechend – wird ebenso angesprochen wie der Aspekt des „Interaktionismus“ mit Einflüssen seitens der Umwelt.

Mandler (1979, 32) dagegen formuliert: **Psyche** ist ein komplexes, einem Individuum zugeschriebenes Informationsverarbeitungssystem, „das Input verarbeitet (einschließlich dem Input aus seinen eigenen Handlungen und Erfahrungen) und Output an die verschiedenen Subsysteme und die Außenwelt abgibt.“ In dieser Umschreibung des Forschungsfeldes der Psychologie wird Mandler sowohl den unbewussten als auch den bewussten Prozessen gerecht, indem er die *Psyche* als komplexes Regulationssystem definiert, innerhalb dessen dem *Bewusstsein* nur eine „Lupenfunktion“ zukommt (s. unten).

Dörner und Selg (1996, 20) definieren im Sinne der Kybernetik: Psychologie ist die „Wissenschaft von den offenen oder variablen **Regulationen**“ (Bischof, 2016). Als „offen“ werden Regulationen dann bezeichnet, wenn sie „nicht genau durch genetische Vorprogrammierungen“ festgelegt sind (Dörner & Selg, 1996, 20). Gemeint sind kybernetische Regelsysteme, die sich plastisch entwickeln können (z.B. Lern- und Denkvorgänge) und nicht genetisch fixiert sind (z.B. Reflexe oder *Erbkoordinationen*). Dass die Unterscheidung zwischen variablen und stabilen Regulationen auf empirischer Basis – zumindest bis heute – noch äußerst schwerfällt, erschwert allerdings die Anwendung dieser Definition.

Dörner und Selg (1996, 24) formulieren weiter: „Gegenstand der Psychologie kann alles werden, was erlebbar ist und / oder sich im Verhalten äußert [...]“. Übereinstimmend mit einigen vorigen Definitionen werden hier introspektives Erleben und beobachtbares Verhalten als gleichwertige Datenquellen der Psychologie verstanden. Vorteilhaft an dieser breiten, aber pragmatischen Definition erscheint außerdem ihre Orientierung in Richtung **Interdisziplinarität** und **Transdisziplinarität**, ohne die eine erschöpfende und realitätsnahe Erklärung psychischer Phänomene kaum möglich ist.

Regulation ist eine Steuerung, welche die Stabilität eines dynamischen Systems aufrechterhält.

Merksatz

Psychologie ist eine Erfahrungswissenschaft, die in möglichst erschöpfender Breite und mit möglichst großer Realitätsnähe die Psyche bzw. ihre „Produkte“ erforscht, nämlich das Verhalten, Erleben und Bewusstsein von Lebewesen.

Interdisziplinarität ist die Zusammenarbeit verschiedener Wissenschaftsdisziplinen zur Lösung eines Problems. Transdisziplinarität erfordert den Einbezug von Praktikerinnen und Praktikern in den wissenschaftlichen Diskurs.

Allgemeine Zielsetzungen wissenschaftlicher Psychologie

| 2.2

In verbreiteten Einführungen der Psychologie (vgl. etwa Bourne & Ekstrand, 2005; Gerrig & Zimbardo, 2008; Ulich, 2000)

Box 2.1 | Häufige Artefakte bei Befragungen

- Unklarheiten in der Formulierung von Fragen (z.B. Mehrdeutigkeit, zu komplizierte Sätze)
- Fehlinterpretationen von Anweisungen („Instruktionen“)
- *Sequenzeffekte* (Ermüdung, „Trainingseffekte“)
- *Hawthorne-Effekt* (sich beobachtet oder analysiert zu fühlen, erhöht zumeist die Leistungsbereitschaft)
- Mangelnde Bereitschaft zur Selbstenthüllung (bei privaten Inhalten)
- Motive zur *Selbstdarstellung*, Effekt der *sozialen Erwünschtheit* (bei Interviewpartnerinnen und -partnern einer Befragung einen bestimmten Eindruck zu hinterlassen, sich nicht zu blamieren etc.)
- Befürchtung negativer Konsequenzen (Zweifel an anonymer Verarbeitung der Daten)
- *Sponsorship-Bias* (Vermutungen über die Absichten der Auftraggeberinnen und -geber von Befragungen)
- *Kontext-Effekte* (z.B. Einfluss von Stimmungen)
- *Urteilsheuristiken* (pragmatische, zeitsparende und oft unlogische Art der Schlussfolgerungen)
- *Anwesenheitseffekte* (Beeinflussung des Antwortverhaltens durch anwesende Personen)

In Anlehnung an Bortz & Döring (1995)

finden sich – gut vergleichbar mit anderen empirischen Sozial- und Humanwissenschaften (wie etwa der Soziologie, der Ökonomie oder der Medizin) – vier Hauptziele für die Wissenschaftsdisziplin Psychologie:

2.2.1 | Beschreiben

Darunter versteht man das (möglichst) präzise, systematische und theoriegeleitete Erfassen von Informationen (*Daten*) über die zu untersuchenden psychischen Phänomene. Häufig verwendete Erhebungsverfahren sind *Selbst- und Fremdbeobachtungen*, *Befragungen*

(Interviews), Experimente, Tests, nichtreaktive Verfahren (z.B. Archive, Abnutzungsgrad von Böden oder Gebrauchsgegenständen), Textanalysen (z.B. Tagebücher), Inhaltsanalysen (Häufigkeit und Bedeutung verwendeter Begriffe), Skalierungen (Semantisches Differential bzw. Polaritätsprofil), Simulationen (z.B. Computermodelle, Szenarien), hirnelektrische Ableitungen (z.B. EEG), Messungen (z.B. Reaktionszeiten) oder Labordaten (z.B. blutchemische Werte). Die Auswahl der Beschreibungsmittel von psychologischen Phänomenen richtet sich primär nach der wissenschaftlichen Grundorientierung der forschenden Person, nach der Art des Phänomens, und bei quantitativen Daten auch nach deren statistischer Verwertbarkeit.

Als **Objektivitätsproblem** bezeichnet man die Schwierigkeit, Daten unverfälscht zu erfassen (Box 2.1). Bei diagnostischen Verfahren zur Beschreibung von Störungsbildern oder Personenmerkmalen werden hohe Gütekriterien gefordert, die sinngemäß für alle psychologischen Datenerhebungen gelten (s. 3.6):

1. **Objektivität:** Sie ist umso größer, je ähnlicher die Daten bei unterschiedlichen datenerhebenden Personen sind.
2. **Reliabilität:** Die sogenannte „Zuverlässigkeit“ von Daten ist umso größer, mit je weniger Erhebungsfehlern sie überlagert sind.
3. **Validität:** Die „Gültigkeit“ von Daten nimmt in dem Maße zu, in dem sie tatsächlich jene Eigenschaft beschreiben, die registriert werden soll (z.B. Intelligenz und nicht auch Konzentration oder Bildung).

Daneben sollten jedoch noch weitere Qualitätsanforderungen an psychologische Daten gestellt werden, nämlich bezüglich der *Skalierung* (Wiedergabe korrekter Quantitäten), der *Normierung* (Normen bzw. Bezugssysteme für Ergebnisse sollen vorhanden sein), der *Fairness* (Daten über verschiedene soziale Gruppen dürfen nicht systematisch verfälscht sein), der *Ökonomie* (der Aufwand der Datenerhebung soll vertretbar sein), der *Zumutbarkeit* (Konsequenzen für Probanden sowie deren Akzeptanz sind zu berücksichtigen), der *Unverfälschbarkeit* (Ergebnisse sollen nicht manipulierbar sein) und der *Nützlichkeit* (Daten sollen zweckentsprechend sein).

Merksatz

Die Beschreibung von Forschungsphänomenen in der Psychologie (Datenerhebung) geschieht hauptsächlich über Selbst- und Fremdbeobachtung, Befragung, Messung, Experiment, Test, Textanalyse, Inhaltsanalyse, Skalierung, Simulation oder Fallstudien, wobei einer verfälschungsfreien Erfassung der Daten besondere Beachtung geschenkt wird (Gütekriterien).

2.2.2 | Erklären

Eine zweite wichtige Zielsetzung der Psychologie ist die Erklärung der beobachteten oder gemessenen Phänomene. Dies geschieht durch **Gesetze** oder durch deren Zusammenfassungen, die **Theorien**. Diese werden durch Ableitung von **Hypothesen** über zu erwartende Ergebnisse in empirischen Untersuchungen getestet. Die Resultate dieser Befragungen, Experimente oder Beobachtungen werden inhaltlich interpretierend (*qualitativ*) oder statistisch (*quantitativ*) auf Gesetzmäßigkeiten überprüft und mit den hypothetisch postulierten Zusammenhängen verglichen. Stimmen die empirisch gefundenen Zusammenhänge mit den erwarteten überein, dann spricht man von einer **Verifikation** der Hypothesen, im gegenteiligen Fall von

deren **Falsifikation**. Eine solche Hypothesentestung setzt die Formulierung einer Theorie oder zumindest die Vorannahme einer Gesetzmäßigkeit voraus. In diesem Falle spricht man von einer **konfirmativen** (bestätigenden) Vorgangsweise, im Gegensatz zu einem **explorativen** Verfahren, wenn es darum geht, an einem Pool gewonnener Daten unbekannte Zusammenhänge erst zu finden.

Merksatz

Hypothesen sind wissenschaftlich begründete Annahmen (Wenn-dann-Aussagen) über Zusammenhänge von Ereignissen. Bestätigte Hypothesen nennt man Gesetze. Als Theorie bezeichnet man zumeist ein System von Gesetzen.

Gesetze und Hypothesen sind zumeist in Form von „Wenn-dann-Aussagen“ formuliert und beziehen sich auf vermutete Kausalzusammenhänge in der Realität. Die „Wenn-Komponente“ von Hypothesen beschreibt jeweils die Ursachen, Bedingungen oder Auslöser von Wirkungen, während die Effekte oder ausgelösten Veränderungen in der „Dann-Komponente“ formuliert werden (Box 2.2; Westermann, 2000). Ein Beispiel eines Gesetzes aus der Kognitionsforschung (*Yerkes-Dodson-Gesetz*): Eine zu hohe oder zu niedrige psychophysiologische Aktivierung (Wenn-Komponente) verringert die Konzentrations-, Denk- und Gedächtnisleistungen (Dann-Komponente).

Merksatz

Wichtige Qualitätskriterien für Gesetze und Theorien sind ihr Grad an Repräsentativität, ihr Realitätsbezug sowie ihre zeitliche und situative Stabilität.

Ein grundlegendes Problem bei der Interpretation von Untersuchungsergebnissen, das sogenannte **Repräsentativitätsproblem**, ist die Frage nach der Verallgemeinerbarkeit, nämlich danach, wie gut von den jeweils beobachteten Daten – den

Fällen der *Stichprobe* – auf die Grundgesamtheit bzw. *Population* zu schließen ist (s. auch Replikationsproblem; Open Science Collaboration, 2015).

Eine andere Unsicherheit besteht darin, ob die abstrakt formulierten Theorien eine inhaltliche Entsprechung in den empirisch ausgewählten Untersuchungsverfahren finden: Die Rede ist vom **Operationalisierungsproblem** bzw. **Validitätsproblem**. Hier geht es etwa darum, ob die *Intelligenz* eines Menschen (d.h. die abstrakte Annahme über die geistige Leistungsfähigkeit einer Person) tatsächlich durch spezielle Intelligenzaufgaben eines Tests erfassbar bzw. ob die theoretische Vorstellung über Intelligenz anhand von anschaulich-konkreten Daten überprüfbar ist.

Weiters ist im sozialwissenschaftlichen Bereich kaum davon auszugehen, dass eine einmal gefundene Gesetzmäßigkeit an allen möglichen Orten, zu allen möglichen Zeiten und unter allen möglichen Umständen gilt, was als **Reliabilitätsproblem** bezeichnet wird (Bortz & Döring, 1995; Schnell, Hill & Esser, 2005). Zur Überprüfung der *Reliabilität* von Ergebnissen bedient man sich verschiedener statistisch gestützter Methoden, bei denen zum Beispiel ein Test für eine psychische Eigenschaft bei gleichen Personen wiederholt eingesetzt wird („Retest-Reliabilität“) oder die Ergebnisse verschiedener Tests zur gleichen Eigenschaft miteinander verglichen werden („Paralleltest-Reliabilität“).

Eine für die Verlässlichkeit von Forschungsergebnissen wichtige Bedingung ist deren Replizierbarkeit bzw. Reproduzierbarkeit, welche in letzter Zeit in mehreren Disziplinen nicht zufriedenstellend ausfiel (Baker, 2016), sodass Verbesserungen im Forschungsprozess vorgeschlagen wurden (Erdfelder & Ulrich, 2018; Fiedler, 2018).

Reliabilität = bedingungsunabhängige Verlässlichkeit einer Datenerhebung

Vorhersagen

2.2.3

Die Formulierung von Gesetzen dient auch zur Erstellung von Prognosen. Wenn zum Beispiel über einen spezifischen Sachverhalt Informationen gegeben sind, dann können unter Verwendung der psychologischen Gesetze Rückschlüsse auf weitere nicht bekannte Merkmale des Sachverhalts gezogen werden (Box 2.2). In der Fachliteratur ist der Fundus an psychologischen Vorhersagen unüberschaubar groß, und die Prognosegüte für zahlreiche Praxis-situationen ist vielversprechend (Frey, Hoyos & Stahlberg, 1988;

Box 2.2 | Prognosen durch Gesetze

Bekannte **Vorinformationen** (Prämissen):

Person X ist sprachbegabt.

Person X ist lernmotiviert.

Person X hat gute Lernbedingungen.

Gesetze (Prämissen):

Wenn eine Person sprachbegabt und lernmotiviert ist sowie gute Lernbedingungen vorfindet, dann erzielt sie höchstwahrscheinlich gute Lernleistungen in Fremdsprachen.

Schlussfolgerung (Konklusion): Person X wird sehr wahrscheinlich gute Lernleistungen in Fremdsprachen erbringen.

Baumann & Perrez, 1990, 1991; Schwarzer, 1997; Hellbrück & Fischer, 1999; Süss & Negri, 2019 usw.): Welche Erziehungsmaßnahmen fördern eine gesunde Entwicklung von Kindern und Jugendlichen? Welche häuslichen Bedingungen sind Voraussetzungen für gute Schulleistungen? Welche Kommunikationsformen erleichtern die berufliche Kooperation? Welche Einflüsse hat die Lebensumwelt auf das Wohlbefinden? Wie kann man am besten Ängsten und Depressionen begegnen? Wie lernt man am schnellsten große Stoffmengen?

Merksatz

Aus psychologischen Gesetzen können vielfältige Vorhersagen über psychische Strukturen oder Abläufe und über deren Abhängigkeit von Umweltbedingungen abgeleitet werden.

Grundsätzlich können Vorhersagen über die Struktur von psychischen Phänomenen (z.B. Intelligenzstruktur, Persönlichkeitsstruktur, Einstellungsprofil) und über deren Dynamik (z.B. Reifungsprozesse, geistige Entwicklung, Entstehung psychischer Störungen) getroffen werden. Ähnlich wie bei politischen Wahlprognosen hängt auch im psychischen Bereich der Erfolg der Vorhersagen wesentlich von der Güte der verwendeten Theorien und der mathematisch-statistischen Prognoseverfahren ab.

Verändern

2.2.4

Eine dauerhafte Veränderung bzw. Optimierung menschlichen Erlebens und Verhaltens (die Veränderung von Gefühlen, Einstellungen, Motiven, Entscheidungen etc.) lässt sich in den meisten Fällen nicht allein durch Vermittlung von Einsichten (z.B. über Kindheitstraumen), durch Anwendung „psychologischer Tricks“ (z.B. *paradoxe Intervention*) oder durch einzelne suggestive Maßnahmen (z.B.

Box 2.3

Arten psychologischer Intervention

- Beobachtungen und Befragungen (haben an sich schon indirekte Auswirkungen, z.B. durch Reflektieren des eigenen Verhaltens oder durch Problematisieren von Befragungsinhalten)
- Kommunikationsstil (kann meinungsbildend, kommunikationsfördernd und konfliktlösend wirken, z.B. durch Maßnahmen der Moderation oder *Mediation*)
- Aufklärung und Bildung (vermittelt psychologisches Wissen und Können, z.B. über optimales Lernen, Möglichkeiten der Stressbewältigung)
- Beratung (Schulberatung, Berufsberatung, Erziehungsberatung, Coaching etc.)
- Training (z.B. Entspannungstechniken, Lerntechniken, Kommunikations- und Kooperationstraining, Elterntaining)
- Therapie (z.B. Kognitive Verhaltenstherapie, Gesprächstherapie)
- Umweltgestaltung und *Partizipation* (z.B. Mitwirkung bei Planungen für menschengerechtes Wohnen und Siedeln, für eine humane Arbeitsplatzgestaltung oder für eine zukunftsfähige Mobilität)

Hypnose) bewerkstelligen, sondern es sind sehr oft komplexe Vorgangsweisen nötig. Dabei müssen nicht nur die Klientinnen und Klienten, sondern auch deren soziale und physische Umfelder einbezogen werden. Der beratenden, pädagogischen oder therapeutischen Anwendung solcher Veränderungsprogramme gehen oft umfangreiche Studien an hunderten Versuchspersonen voraus, um den Erfolg unter möglichst vielen Bedingungen sicherzustellen.

Die Liste möglicher psychologischer Einflussnahmen ist relativ groß und beginnt schon damit, dass Personen sich anders verhalten, wenn sie sich beobachtet fühlen. Ein gutes Beispiel dafür ist der sogenannte **Hawthorne-Effekt**, nach den amerikanischen „Western Electric Hawthorne Works“ in Chicago benannt, einer Fabrik, in der in den Jahren 1924 bis 1927 Elton Mayo den Einfluss von Arbeitsbedingungen auf die Produktivität testete: Er kam zum Schluss, dass man mehr zu leisten bereit ist, wenn man sich (z.B. im Zuge einer wissenschaftlichen Untersuchung) beobachtet fühlt (Flick et al., 1995).

Der Kommunikationsstil nimmt in mehrfacher Weise Einfluss: Kommunikationspsychologisch geschulte Moderatorinnen und Moderatoren können in Diskussionsrunden den Meinungsaustausch und die Konfliktbewältigung wesentlich dadurch fördern, dass sie eine partnerschaftliche Atmosphäre mit fairen Regeln für die Argumentation schaffen (Montada & Kals, 2013). Durch Kommunikationstechniken – wie dem „Partnerzentrierten Gespräch“, dem „Kontrollierten Dialog“ und der „Themenzentrierten Interaktion“ – sinkt in Partnerschaften, Arbeitsgemeinschaften und Firmen die Streithäufigkeit, während die Kooperationsfähigkeit steigt.

Bereits wesentlich aufwendiger gestaltet sich der Einsatz psychologischen Wissens für Beratungstätigkeiten im Bildungs-, Arbeits- und Therapiebereich. Ähnlich wie in der Medizin werden hier vorerst die jeweiligen Ausgangsbedingungen erhoben (*Anamnese*) und die Probleme und Störungen festgestellt bzw. analysiert (*Diagnose*). In manchen Fällen sind umfangreiche Testungen, wie etwa zur Feststellung der Begabungsorientierung, der Interessenausrichtung, des Motivationsprofils oder der allgemeinen Problemsituation der Klientinnen und Klienten, nötig.

Besonders spezialisiert und auf die Art und Bedingtheit der behandelten Störung maßgeschneidert (s. Reinecker, 2003a) sind die in der empirischen Psychologie entwickelten Therapieverfahren (insbesondere Verhaltenstherapie, Gesprächstherapie, Familientherapie). Sie sind im Rahmen eines Psychologiestudiums aufgrund des hohen Übungsbedarfs und der nötigen Supervision nicht ausreichend lern- und trainierbar und müssen daher in anspruchsvollen

Merksatz

Auf Basis psychologischer Gesetze und Theorien konnte eine große Vielfalt von Maßnahmen (Interventionen) zur Veränderung problematischen Verhaltens, Erlebens und Bewusstseins entwickelt werden.

Zusatzausbildungen nach dem Studium vermittelt werden (z.B. in Österreich die postgraduale Ausbildung für „Klinische und Gesundheitspsychologie“ und / oder für „Psychotherapie“).

Weitere nicht unwichtige Einflussmöglichkeiten der Psychologie liegen im Bereich der Evaluation (Wottawa & Thierau, 2003) und Intervention im Wohn-, Wirtschafts-, Arbeits- und Bildungsbereich. Über die sogenannte „User Needs Analysis“ (UNA), „Post Occupancy Evaluation“ (POE) oder „Environmental Impact Analysis“ (EIA) lassen sich zum Beispiel wichtige Lebensbedürfnisse des Menschen ermitteln und Vorschläge für deren Befriedigung erarbeiten sowie eine allgemeine Verbesserung der Lebensumstände schaffen (Harloff, 1993). In neuerer Zeit werden immer mehr moderne Technologien auf ihre psychologische Nutzbarkeit hin untersucht (EDV-Arbeitsplätze, Internet-Aktivitäten, E-Learning, Teleworking etc.).

Kontroversielle Grundannahmen der Psychologie

| 2.3

Seit etwa 1960 hat sich an den universitären Psychologieinstituten im deutschen Sprachraum eine „Mainstream-Psychologie“ durchgesetzt, nämlich jene mit naturwissenschaftlicher, empirisch-statistischer Orientierung. In den Achtzigerjahren meinte Hofstätter (1984, 103): „Die Konflikte zwischen den Richtungen und Schulen gehören fast überall in der Psychologie der Vergangenheit an“, und begründete dies damit, dass kaum mehr der Anspruch erhoben werde, mit „gleichen Prinzipien die verschiedenen Problemfelder“ der Psychologie aufzuklären. Tatsächlich ist seit den späten Sechzigerjahren die Heftigkeit der Auseinandersetzung zwischen den verschiedenen wissenschaftlichen Strömungen innerhalb der Psychologie erheblich zurückgegangen, was aber nicht gleichzeitig bedeutet, dass die wechselseitige Akzeptanz zugenommen hätte. Immer noch bestehen zwischen Angehörigen des Faches – vor allem aber zwischen Praktikern und Forschern – erhebliche Meinungsunterschiede darüber, was unter wissenschaftlichem Vorgehen zu verstehen ist, welche Themen als forschungswürdig anzusehen sind und was als allgemein verbindlicher Wissensbestand der Psychologie zu betrachten ist.

Auch wenn sich Wissenschaftlerinnen und Wissenschaftler zu meist nicht explizit zu ihren inhaltlichen oder methodischen Grundannahmen (Forschungsaxiomen) bekennen, so lassen sich

3.1 | Wissenschaftlichkeit

Wissenschaftliches Handeln sollte sich an logisch begründeten, explizit formulierten und verbindlichen Kriterien orientieren. Nach Wohlgenannt (1969) sowie Konegen und Sondergeld (1985) sind dies folgende:

- Es sollen nur Aussagen über Sachverhalte gemacht werden, die wirklich vorhanden sind (Beobachtbarkeit bzw. Erlebbarkeit).
 - Die Aussagen sollen ein System bilden und nach expliziten (wissenschaftsspezifischen) Regeln zustande kommen.
 - Es müssen Regeln zur Definition von Fachausdrücken (Termini) vorhanden sein.
 - Für das gegebene System von Aussagen müssen Ableitungsregeln gelten („induktive“ und „deduktive“ Schlussregeln).
- Das Aussagensystem muss widerspruchsfrei sein.
- Aussagensysteme mit empirischem Bezug (faktische Aussagen) dürfen sich nicht auf die Aufzählung von Fakten beschränken, sondern müssen auch Verallgemeinerungen enthalten.
- Faktische Aussagen müssen intersubjektiv prüfbar sein.

In ähnlicher Weise charakterisieren Bortz und Döring (1995, 7) aus Sicht der Psychologie wissenschaftliche Aussagen: „Wissenschaftliche Hypothesen sind Annahmen über reale Sachverhalte (empirischer Gehalt, empirische Untersuchbarkeit) in Form von Konditionalsätzen. Sie weisen über den Einzelfall hinaus (Generalisierbarkeit, Allgemeinheitsgrad) und sind durch Erfahrungsdaten widerlegbar (Falsifizierbarkeit).“

Merksatz

Wissenschaftliches Vorgehen will für Tatsachen (Fakten) ein möglichst widerspruchsfreies System von mehr oder weniger abstrakten, logisch verknüpften und intersubjektiv prüfbaren Aussagen bilden.

Merksatz

Die Methoden einer wissenschaftlichen Disziplin sollen die korrekte und zweckmäßige „Abbildung“ eines empirischen (konkreten) Systems in einem theoretischen (abstrakten) System erlauben.

3.2 | Von der Empirie zur Theorie

Die Human- und Sozialwissenschaft Psychologie unterscheidet sich insofern grundlegend von den Naturwissenschaften, als hier die wissenschaftlichen Phänomene nicht direkt zugänglich sind, sondern oft indirekt erschlossen werden müssen. So sind etwa Persönlichkeit, Intelligenz oder Einstellungen theoretische Be-

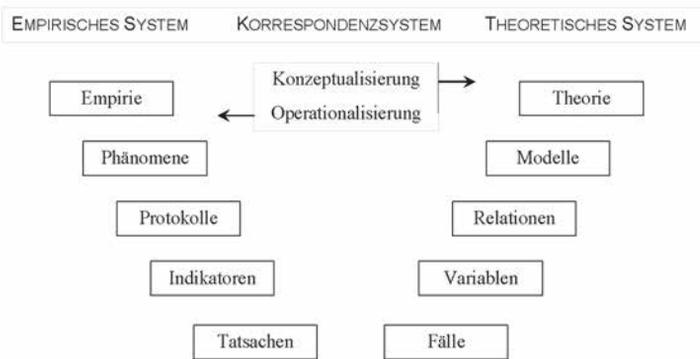


Abb 3.1

Die theoretische Beschreibung der Realität kann als deren abstrakte Abbildung in einem Symbolsystem (Sprache, Vorstellung, Programme ...) aufgefasst werden. Dabei wird ein vermittelndes, transformierendes Korrespondenzsystem benötigt, wodurch wissenschaftscharakteristische Vorschriften zur Gewinnung, Beschreibung, Erklärung und Interpretation der jeweiligen Systemelemente zur Verfügung gestellt werden.

griffe, die nur über Verhaltenstendenzen, Fertigkeiten oder Gefühlsreaktionen erfasst werden können. Der Weg von der Empirie zur Theorie ist daher in der Psychologie oft weit und erfordert viele Zwischenschritte. In der psychologischen Methodenlehre unterscheidet man zumeist ein **empirisches System**, das die Forschungsdaten liefert, und ein **theoretisches System**, das die Gesetze und Erklärungen zu formulieren gestattet, und bezeichnet die Vorgangsweisen, Methoden und Instrumente, die zwischen beiden eine Verbindung herstellen, als **Korrespondenzsystem**.

Die Schritte vom empirischen zum theoretischen System (Abb. 3.1) lassen sich wie folgt charakterisieren: Vorerst werden aus einer Vielzahl von Strukturen und Abläufen in der psychischen oder sozialen Realität – der sogenannten **Empirie** – jene Phänomene identifiziert, die Gegenstand der wissenschaftlichen Untersuchungen werden sollen. Die jeweils zu erforschenden **Phänomene** (z.B. Denkprozesse, Lernformen, Stressverarbeitung) müssen exakt beschrieben werden, was sich in verbalen, bildlichen oder symbolischen Datenmengen bzw.

Merksatz

Im empirischen System werden Phänomene menschlicher Erfahrung ausgewählt und in ihren konkreten Erscheinungsweisen (Tatsachen) verbal oder symbolisch (über Indikatoren) protokolliert.

Protokollen niederschlägt. Insbesondere in der quantitativen Forschung versucht man die Datenmenge auf jene Informationseinheiten zu begrenzen, die zur Beschreibung der Gesetzmäßigkeiten der Phänomene relevant erscheinen. Die Gesamtheit aller Ausprägungen von **Indikatoren** (zu einem untersuchten Phänomen) zu einer bestimmten Zeit und an einem bestimmten Ort sind jene einzelnen **Tatsachen**, die der empirischen Forschung als Grundeinheiten zur Gewinnung oder Überprüfung von (statistischen) *Hypothesen* zur Verfügung stehen.

Merksatz

Im theoretischen System werden anhand von Fällen Relationen zwischen Variablen gesucht, funktional zusammenhängende Relationen zu Modellen zusammengefasst und thematisch verwandte Modelle zu einer Theorie integriert.

Zu einem **Fall** des theoretischen Systems wird eine Tatsache dann, wenn die erfassten Eigenschaften der Indikatoren in Ausprägungen von **Variablen** umgewandelt werden. Ein Fall ist somit durch eine bestimmte Konfiguration von (empirischen) Variablen, genauer durch deren jeweilige Ausprägungen, definiert. Mittels statistischer Auswertungsverfahren werden auf Basis der zur Verfügung stehenden Fälle zwischen den Variablen entweder hypothetische **Relationen** (Funktionen,

Beziehungen etc.) geprüft oder unbekannte Relationen gesucht. **Hypothesen** sind Annahmen über *Relationen* zwischen mindestens zwei (empirischen) Variablen. Wenn eine hypothetische Relation zufriedenstellend oft in verschiedenen, wissenschaftlich seriösen Untersuchungen empirisch bestätigt wurde, spricht man von einem **Gesetz**. Mehrere Gesetze (oder Hypothesen), die ein logisch konsistentes Erklärungsgerüst für ein bestimmtes Phänomen darstellen, werden zusammenfassend als **Modell** bezeichnet (z.B. Wahrnehmungs-, Lern-, Gedächtnis- und Handlungsmodelle). Der Übergang von Modellen zu Theorien ist fließend. Eine **Theorie** ist ein System von zusammenhängenden Gesetzen, die maximal abstrakt formuliert sind.

Merksatz

Konstrukte sind speziell definierte, nicht direkt beobachtbare Begriffe einer psychologischen Theorie (z.B. Intelligenz, Motivation, Aggression), für die Operationalisierungen vorhanden sind oder entwickelt werden müssen.

Um die Realität in Form von Gesetzen oder Theorien abbilden zu können, müssen *Begriffe* (*Konzepte*) zur Klassifikation empirischer Phänomene entweder vorhanden sein (Alltagsbegriffe) oder neu entwickelt werden (Fachbegriffe bzw. Ter-

mini). Diese **Konzeptionalisierung** („Konzeptspezifikation“; Schnell, Hill & Esser, 1993) der Wahrnehmungs- oder Erlebenswelt darf weder zu fein noch zu grob ausfallen, damit ein adäquater Auflösungsgrad für die untersuchten Phänomene gegeben ist. Für neu eingeführte theoretische Fachbegriffe, sogenannte **Konstrukte** (d.h. theoretische Konstruktionen), ist die konkrete Bedeutung in der Welt unserer Erfahrungen mittels **Operationalisierungen** klarzulegen. Als solche Interpretationshilfen für theoretische Fachbegriffe können spezielle Beobachtungen, Testverfahren, Teile von Fragebögen oder sonstige Datenerfassungsverfahren herangezogen werden. Mögliche Operationalisierungen von „Angst“ sind etwa bei einem Versuchstier der körperliche Zustand in Erwartung elektrischer Schläge, die gemessene Herzfrequenz oder die motorische Unruhe. „Intelligenz“ kann durch die Leistungen in einem bestimmten Intelligenztest, und „Glück“ durch die Beantwortung von Fragen in einem Befindlichkeitstest operationalisiert werden. Das Korrespondenzsystem mit einschlägigen Konzeptualisierungen und Operationalisierungen ist Bestandteil des jeweiligen *wissenschaftlichen Paradigmas* (s. auch Maderthaner, 2003).

Merksatz

Bedeutungsinterpretierende Zuordnung beobachtbarer Sachverhalte zu einem theoretischen Begriff bezeichnet man als dessen Operationalisierung.

Fälle und Variablen

3.3

Als **empirische Einheiten** kommen in der Psychologie beliebige statische oder dynamische Systeme infrage (z.B. Personen, Gruppen, Situationen, Abläufe), in denen sich psychische Gesetzmäßigkeiten äußern. Wie bereits erwähnt, wird die Beschreibung (*Protokoll*) eines *Phänomens* auf gesetzesrelevante Merkmale (*Indikatoren*) reduziert, sodass zuletzt nur mehr ein sogenannter „Fall“ mit phänomencharakteristischen Variablen übrig bleibt. *Fälle* sind also die – im Sinne einer wissenschaftlichen Fragestellung – maximal informationsreduzierten empirischen Einheiten, anhand derer Gesetze verifiziert oder falsifiziert werden sollen.

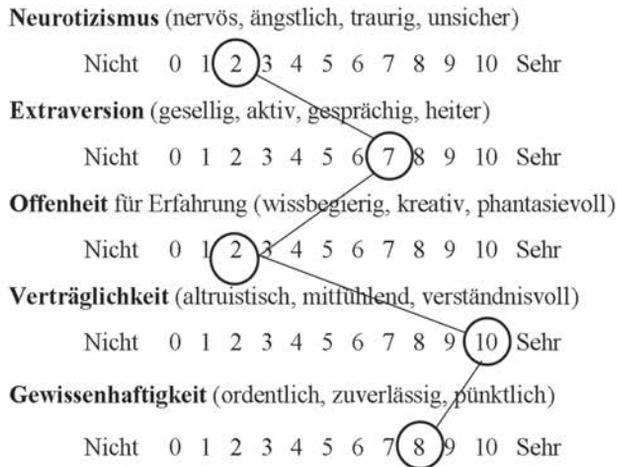
Da es in der Psychologie nur selten möglich ist, die gesamte **Population** bzw.

Merksatz

In sozialwissenschaftlichen Untersuchungen werden anhand von Fällen (Stichprobe) Gesetze gewonnen, welche auf ähnliche Sachverhalte (Population, Geltungsbereich der Gesetze) hin verallgemeinert werden.

Grundgesamtheit empirischer Einheiten zu erfassen, für die ein Gesetz gelten soll, beschränkt man sich in der Forschung auf eine **Stichprobe** (engl. sample), deren Zusammensetzung in den gesetzrelevanten Eigenschaften jener der Population möglichst ähnlich sein sollte, damit die auf Basis der Stichprobe gewonnenen Erkenntnisse berechtigt verallgemeinert werden können. Der Schluss von der Stichprobe auf die Population ist am ehesten dann gerechtfertigt, wenn die Stichprobe nach dem Zufallsprinzip aus der Grundgesamtheit ausgewählt wird (*Randomisierung*) und die Stichprobe entsprechend groß ist (s. auch Schnell et al., 1993).

Abb 3.2



Ein Beispiel für eine einfache, aber prägnante Charakterisierung von Personen (Fällen) ist jene nach Persönlichkeitsfaktoren (Variablen). Das Profil in der Abbildung kennzeichnet eine Person in den sogenannten „Big-Five-Faktoren“ („NEO Five-Factor Inventory“ von Costa & McGrae, 1992; Becker, 2004).

Wenn die Ausprägungen relevanter Untersuchungsvariablen in einer Stichprobe mit jenen der Population annähernd übereinstimmen, darf von **Repräsentativität** der Stichprobe gesprochen werden. Im Forschungsalltag ist Repräsentativität aufgrund verschiedenster Forschungshemmnisse nur selten vollständig erreichbar (Kostenbegrenzung, Unerreichbarkeit von Personen, Teilnahmeverweigerung etc.), sodass häufig nur **Gelegenheitsstichproben** (z.B. Studierendensamples) zur Verfügung stehen oder die Stichprobenselektion eher mittels **Quotaverfahren** (Vergleichbarkeit der Stichprobe

mit der Population hinsichtlich der Verteilung einiger wichtiger Merkmale wie Geschlecht, Bildung, Beruf usw.), mittels **Schneeballverfahren** (Probandinnen und Probanden vermitteln selbst wieder weitere Probandinnen und Probanden) oder mittels **Klumpenverfahren** („cluster sampling“: Cluster von Fällen, z.B. Unternehmen, Organisationen, Branchen, werden zufällig ausgewählt und hierin alle Mitglieder untersucht). Leider erhöhen die letztgenannten Auswahlverfahren die Fehleranfälligkeit und mindern den Grad an Verallgemeinerbarkeit.

In der Mathematik sind **Variablen** („Platzhalter“, „Leerstellen“) jene Zeichen in Formeln, die für einzelne Elemente aus einer Menge möglicher Zahlen oder Symbole stehen. Die verschiedenen Belegungen von Variablen nennt man ihre **Ausprägungen** oder – wenn diese aus Zahlen bestehen – ihre **Werte**. Als **Wertebereich** einer Variablen bezeichnet man alle Zahlen vom Minimal- bis zum Maximalwert. Variablen charakterisieren Fälle hinsichtlich ihrer untersuchungsrelevanten Merkmale. In psychologischen Untersuchungen können diese äußerst vielfältig sein und schließen Beschreibungsmerkmale, Testergebnisse, Prozentschätzungen, physiologische Messwerte und andere Aspekte mit ein (Abb. 3.2).

Während in der Mathematik Zahlen definitionsgemäß eine quantitative Bedeutung haben, das heißt, dass bestimmte Rechenoperationen mit ihnen durchgeführt werden können (Addition, Multiplikation, Potenzierung etc.), kann dies bei Variablenwerten der psychologischen Empirie nicht vorausgesetzt werden. Hier können Zahlen zum Beispiel für Benennungen herangezogen werden (z.B. Abzählung von Personen in einer Gruppe), sie können eine Rangordnung symbolisieren (z.B. der 1., 2. oder 3. in einem Wettkampf) oder sie können ein Vielfaches von Grundeinheiten darstellen (z.B. Häufigkeiten). Aus diesem Grund werden die Ausprägungen von Variablen in der Psychologie hinsichtlich ihrer sogenannten *Skalenqualität* unterschieden, wovon insbesondere die Anwendbarkeit statistischer Auswertungsverfahren abhängt.

Faktoren, denen innerhalb von Phänomenen ein Einfluss zugeschrieben wird, heißen in den empirischen Sozialwissenschaften (so wie in der Mathematik bei Funktionsgleichungen) **unabhängige Variablen (UV)**, während jene Faktoren, welche die Auswirkungen des Einflusses symbo-

Merksatz

Hypothetische Ursachen werden in empirischen Untersuchungen mittels unabhängiger Variablen charakterisiert und hypothetische Wirkungen mittels abhängiger Variablen.

lisieren, als **abhängige Variablen (AV)** bezeichnet werden. In einer wissenschaftlichen *Kausalhypothese* (s. 3.4) stellt der Wenn-Teil die Ausprägungen der unabhängigen Variablen und der Dann-Teil die vorhergesagten Ausprägungen der abhängigen Variablen dar (Box 3.1). Auf diese Unterscheidung verzichtet man, wenn die Einflussrichtung zwischen den Variablen nicht spezifiziert ist oder als wechselseitig angenommen wird (z.B. bei *Korrelationsstudien*).

Von den eigentlichen Wirkvariablen unterscheidet man sogenannte **Moderatorvariablen**, denen ein modifizierender Einfluss auf die funktionalen Beziehungen zwischen unabhängigen und abhängigen Variablen zugeschrieben wird (Box 3.1).

Da selbst bei bestens geplanten und genau kontrollierten Experimenten Einflüsse wirksam werden, die nicht erwünscht sind, existieren in allen empirischen Untersuchungen auch **Störvariablen**. Je mehr Störeinflüsse in einer Untersuchung vorhanden sind, desto vager und unschärfer werden die wissenschaftlichen Resultate. Von **Konfundierung** spricht man, wenn unabhängige Variablen mit dem Effekt anderer Variablen vermischt sind. Gebräuchliche Maßnahmen gegen eine Verfälschung durch konfundierende Variablen oder Störvariablen sind deren

- „Elimination“ (d.h. Versuch ihrer Ausschaltung),
- „Matching“ (d.h. Gleichhaltung ihres Effektes bei den Ausprägungen der unabhängigen und abhängigen Variablen) sowie
- *Randomisierung* (d.h. zufällige Aufteilung ihrer Quellen, wie etwa der Auswahl der *Probandinnen und Probanden*).

Insbesondere bei modernen statistischen Modellen findet man häufig die Unterscheidung in **manifeste** und **latente Variablen**. Als manifest gelten alle durch direkte Erhebung (als Ergebnis der empirischen Datenerhebung) zustande gekommenen Variablen, während latente Variablen theoretisch begründet sind und zur Erklärung der empirischen Resultate herangezogen werden. So etwa kann das *Konstrukt* Intelligenz durch eine latente Variable beschrieben werden, wenn diese als Summe aller gelösten Intelligenzaufgaben definiert wird. Die Ausprägungen latenter Variablen werden in der Forschungspraxis mittels mehr oder weniger komplexer mathematischer Prozeduren (z.B. über *Mittelwertbildungen*, *lineare Funktionen*, *probabilistische Schätzungen*) aus den ihnen über die Operationalisierung zugeordneten manifesten Variablen errechnet.

lat. confundere:
zusammengießen,
vermischen, vermengen,
verwirren

probabilistisch: wahr-
scheinlichkeitstheore-
tisch berechnet

Beispiel für eine Variablentypisierung

| Box 3.1

Wenn etwa in einem Experiment der Einfluss des Alkoholkonsums auf die Fahrleistung in einem Fahrsimulator untersucht werden soll, dann könnte die Hypothese lauten: Wenn Verkehrsteilnehmer Alkohol trinken, dann begehen sie überdurchschnittlich viele Fehler im Simulator. Als *unabhängige Variable* fungiert der Alkoholgehalt des Blutes, welcher zumindest in zwei Ausprägungen vorliegen muss (z.B. 0,0 Promille Blutalkoholgehalt – 0,5 Promille Blutalkoholgehalt). Als *abhängige Variable* könnte in einer normierten Fahrleistungsprüfung die Anzahl an Fahrfehlern herangezogen werden. Als *Moderatorvariablen*, welche die Beziehung zwischen Alkoholisierungsgrad und Fahrleistung verändern könnten, wären die Fahrpraxis, die Alkoholtoleranz oder die Trinkgeschwindigkeit der Versuchspersonen einzubeziehen. Als *Störvariablen* können Messfehler bei der Blutalkoholbestimmung, Konzentrationschwankungen der Probandinnen und Probanden oder Ablenkungen in der Versuchssituation angenommen werden.

Kausalität und Wahrscheinlichkeit

| 3.4

Die Annahme, dass Ereignisse der Realität einander gesetzmäßig beeinflussen, d.h. in einem Kausalzusammenhang zueinander stehen, wird implizit in jeder Wissenschaft vorausgesetzt. Würde die Welt nicht deterministischen oder zumindest probabilistischen Gesetzen (wie z.B. in der Quantenphysik) unterliegen, hätte das Betreiben von Wissenschaft keinen Sinn. Das **Kausal(itäts)prinzip**, nämlich die Annahme, dass jedes Ereignis eine oder mehrere Ursachen hat, ist eine grundsätzlich unbeweisbare These, die aber sowohl im Alltag als auch in der Wissenschaft dazu motiviert, immer wieder nach Ursachen und Wirkungen zu fragen. Im Vergleich zur oft trivial vereinfachten Kausalanalyse des täglichen Lebens (z.B.: Wer ist schuld an einer Scheidung? Was ist die Ursache eines Unfalls?) unterscheidet man in der Wissenschaft mehrere Arten von Kausalbeziehungen.

Box 3.3 | Beispiel für eine multivariate aussagenlogische Beschreibung von Variablenzusammenhängen (mit Multikausalität und bedingter Kausalität)

Inhaltliche Aussage (aus der Sozialpsychologie):

Kinder, die früher an Modellpersonen beobachten konnten, dass sich Aggression „lohnt“ (L), oder solche, die gerade von einem anderen Kind frustriert (F) wurden, tendieren diesem gegenüber zu aggressivem Verhalten (A), wenn dieses eher als wehrlos (W) empfunden wird, wenn ihm gegenüber keine moralischen Hemmungen (M) bestehen und wenn im Moment keine Strafdrohung (S) von Aufsichtspersonen für aggressive Reaktionen zu erwarten ist.

Aussagenlogische Form: $((L \vee F) \wedge W \wedge \neg M \wedge \neg S) \rightarrow A$

(Zur Bedeutung der Symbole siehe Abb. 3.6; \neg : Negation)

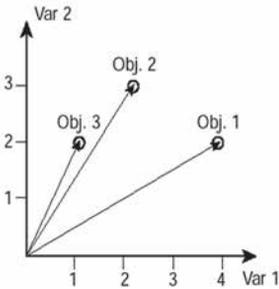
Als eines der größten methodischen Probleme der gegenwärtigen psychologischen Forschung kann gelten, dass die meisten Erklärungsmodelle und Hypothesen weder Multikausalität noch bedingte Kausalität einbeziehen, und daher zu geringe Prognosesicherheit erreichen (Maderthaner, in Vorbereitung).

3.6 | Beschreibende und hypothesenprüfende Statistik

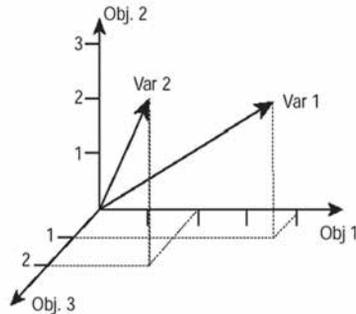
Die **Statistik** (engl. statistics) fungiert innerhalb der Psychologie als Hilfswissenschaft zur Auffindung und Beschreibung von nichtdeterministischen Gesetzen (Relationen). “Statistics is a set of concepts, rules and methods for (1) collecting data, (2) analyzing data, and (3) drawing conclusions from data” (Iversen & Gergen, 1997, 4). Wenn in der Nachrichtentechnik damit gerechnet wird, dass elektronische Signale von Störungen überlagert werden, dann sendet man gleiche Signale mehrmals hintereinander, um beim Empfang auf Basis ihres *Mittelwerts* (Durchschnitt) auf das ursprüngliche „wahre“ Signal schließen zu können. Nach dem gleichen Prinzip

| | Variable 1 | Variable 2 |
|----------|------------|------------|
| Objekt 1 | 4 | 2 |
| Objekt 2 | 2 | 3 |
| Objekt 3 | 1 | 2 |

Abb 3.7



Variablenraum
(Vergleich von Objekten)



Objektraum
(Vergleich von Variablen)

Ähnlich wie physikalische Objekte in einem (euklidischen) Raum Positionen einnehmen können, lassen sich auch Fälle als „Datenobjekte“ auffassen und in einen (multidimensionalen) Variablenraum projizieren. Die Datenobjekte sind einander umso näher, je ähnlicher ihre Variablenausprägungen sind. Ebenso lässt sich die Ähnlichkeit von Variablen in Objekträumen abbilden. Im Beispiel sind drei Objekte (z.B. Personen) im Zweivariablenraum (z.B. Gewicht und Größe) dargestellt und daneben die gleichen Variablen im Dreiobjektraum.

werden in der Psychologie wiederholt oder simultan Daten über psychische Abläufe, Einstellungen oder Fähigkeiten gesammelt, um daraus Schätzungen über die untersuchten Phänomene ableiten zu können. Die (klassische) **Testtheorie** postuliert diesem Prinzip gemäß, dass sich jeder **Messwert** (z.B. eine physiologische Ableitung, eine Fragebogenantwort, eine Prozentschätzung) aus einem **wahren Wert** und einem zufälligen **Fehlerwert** zusammensetzt und dass sich Fehlerüberlagerungen durch Heranziehung mehrerer Messwerte des gleichen Ereignisses „ausmitteln“ lassen.

Merksatz

Mittels der Statistik als Hilfswissenschaft werden in der Psychologie verfügbare Daten beschrieben und auf vorhandene Gesetzmäßigkeiten untersucht.

3.6.1 | Deskriptivstatistik – beschreibende Statistik

Die deskriptive Statistik bietet charakteristische, formelhafte Beschreibungen oder grafische Darstellungen für eine große Zahl von Fällen, Variablen oder Variablenrelationen an. Wie schon erwähnt (Abb. 3.7), werden Fälle als Punkte in einem multidimensionalen Raum von Variablen gedacht, und ihre Verteilung wird durch **statistische Kennwerte** bzw. **Statistiken** näher charakterisiert. Eine Voraussetzung für eine solche Darstellung von Variablen (sowie ihres Einbezugs in komplexe statistische Auswertungsverfahren) ist, dass sie quantitativ interpretierbar sind, d.h., dass ihre Ausprägungen unterschiedliche Quantitäten einer Eigenschaft oder eines Merkmals von Fällen kennzeichnen (Backhaus et al., 2003).

Um die quantitative Bedeutung der Ausprägungen von Variablen einzustufen, werden diese hinsichtlich ihrer **Skalenqualität**, d.h. nach **Skalenniveau** bzw. **Messniveau**, differenziert:

1. **Nominalskala:** Wenn eine Variable nur dieses Skalenniveau zugeschrieben bekommt, sind ihre Ausprägungen (Zahlenwerte) im Sinne von Klassifikationen zu verstehen. Es handelt sich also um Variablen, die („qualitative“) Eigenschaften, wie etwa Geschlecht, Beruf, Nationalität oder Haarfarbe, kennzeichnen.
2. **Ordinalskala (Rangskala):** Die Werte von Variablen mit ordinaler Skalenqualität gestatten nicht mehr bloß die Unterscheidung zwischen gleich- und ungleichartig, sondern erlauben zusätzlich die Erstellung einer quantitativ begründeten Rangreihe der Variablenausprägungen. Typische Ordinalvariablen sind Listen von Schulnoten oder Rangreihungen bei Wettbewerben.
3. **Intervallskala:** Für Variablen dieses Typs wird angenommen, dass ihre aufeinanderfolgenden Zahlenwerte die Zunahme einer variablen-spezifischen Eigenschaftsquantität immer um den gleichen Betrag symbolisieren (der Quantitätszuwachs von 1 auf 2 ist der gleiche wie etwa von 4 auf 5). In der Psychologie erwartet man zumindest Intervallskalenniveau von all jenen Variablen, die quantitative Abstufungen von individuellen Leistungspotenzialen (z.B. Konzentration, Intelligenz) oder von psychischen Dis-

Merksatz

Die quantitative Interpretierbarkeit von empirischen Variablen bzw. der durch sie beschriebenen Indikatorausprägungen wird durch die ihnen zugeschriebene Skalenqualität (Messniveau) charakterisiert.

positionen (z.B. Einstellungen, Persönlichkeitsdimensionen) zum Ausdruck bringen wollen.

4. **Verhältnisskala (Rationalskala):** Variablen dieser Art sind gewissermaßen Intervallskalen mit einem fixen Nullpunkt. In der Psychologie gehören Verhältnisschätzungen für Wahrnehmungsreize diesem Skalentyp an oder bestimmte probabilistische Testkennwerte („Item-Response-Modelle“).
5. **Absolutskala:** So bezeichnet man Variablentypen, die ebenfalls einen fixen Nullpunkt haben, bei denen aber auch „echte“ Einheiten gegeben sind. Zu diesem Typus zählen alle Variablen, die Häufigkeiten bzw. Frequenzen zum Ausdruck bringen (z.B. Schätzungen der Anzahl von Objekten oder Personen).

Als quantitative Variablen im engeren Sinne zählen für die Statistik nur solche, die als Intervall-, Verhältnis- oder Absolutskalen zu interpretieren sind.

Eine empirische Variable hat noch eine weitere, für komplexe statistische Auswertungen wichtige Eigenschaft: die **Verteilung** ihrer Ausprägungen.

Die Betrachtung der Verteilung empirischer Variablen ist aus verschiedenen Gründen wichtig:

1. Aus ihr geht hervor, welche Zahlenwerte mit welcher Wahrscheinlichkeit in einer Population zu erwarten sind (z.B. Mess- oder Testergebnisse).
2. Sie kann Hinweise darüber geben, ob der Wertebereich einer Variablen für die Beschreibung eines empirischen Prozesses optimal gewählt wurde (z.B. nicht optimal bei „schiefen“ Verteilungen, wenn sich die Werte bei den Minimal- oder Maximalwerten der Variablen häufen).
3. Ein weiterer Grund für die **Verteilungsprüfung** von Variablen liegt in der notwendigen Prüfung von Verteilungsvoraussetzungen (z.B. dem Erfordernis der Normalverteilung von Fehlerkomponenten) für bestimmte multivariate statistische Auswertungsmethoden (z.B. der „Regressionsanalyse“).

Die besondere Bedeutung der **Normalverteilung** (oder „Gauß’schen Glockenkurve“) und der (mit ihr verwandten) **Binomialverteilung** in der Statistik ist darauf zurückzuführen, dass beide als Idealformen zufallsbedingter Verteilungsprozesse angesehen werden. Wie bereits erwähnt, wird bei empirischen Variablen angenommen, dass sich ihre Werte aus einer wahren Komponente und einer zufälligen Fehlerkomponente zusammensetzen.

Um den „Schwerpunkt“ von mehreren Variablenwerten zu bestimmen, werden Maße der „Zentraltendenz“ („Lageorientierung“) herangezogen, wie etwa der **Mittelwert** (m) bzw. „Durchschnittswert“, nämlich die Summe (Σ) aller Werte (x) dividiert durch die Anzahl der Werte (n):

$$m = \frac{1}{n} \cdot \sum x$$

Ebenfalls als Maß der Zentraltendenz gebräuchlich ist der **Median** (jener Wert, von dem aus etwa 50 % aller Werte größer oder kleiner sind).

Eine zweite wichtige Kennzeichnung von Variablen sind statistische Kennwerte, die den „Streubereich“ („Dispersion“) der Ausprägungen von Variablen aufzeigen. Maße dafür sind etwa die **Streubreite** (Bereich vom maximalen Wert bis zum minimalen Wert), die **Varianz**

$$v = \frac{1}{n} \cdot \sum (x - m)^2$$

oder die **Standardabweichung**:

$$s = \sqrt{v}$$

Die Varianz (v) ist als Durchschnittswert für die Abweichungsquadrate aller Werte (x) vom Mittelwert (m) definiert. Die Standardabweichung (s) als Wurzel der Varianz bezeichnet jene Abweichungen vom Zentrum der Normalverteilung, innerhalb derer etwa 68 % aller Werte liegen.

Um nun Variablen (mit verschiedenen Mittelwerten und Streuungen) besser miteinander vergleichen und auf wechselseitigen Zusammenhang (*Korrelation*) überprüfen zu können, werden sie oft durch einen einfachen Rechenvorgang in sogenannte **Standardvariablen** mit jeweils einem Mittelwert von 0,0 und einer Streuung von 1,0 umgewandelt (*Standardisierung*). Dies wird erreicht, indem alle Werte einer Variablen um ihren Mittelwert reduziert und durch die Streuung dividiert werden („lineare Transformation“, s. auch Abb. 8.19):

$$z = \frac{x - m}{s}$$

Um das Ausmaß der linearen „Ko-Relation“ zwischen zwei Variablen abschätzen zu können, bedient man sich seit etwa hundert Jahren des Pearson’schen **Korrelationskoeffizienten** (r), der bei einem maximal positiven Zusammenhang zwischen den zwei Variablen den Wert $+1,0$ annimmt (wenn beide gleichsinnig zu- oder abnehmen). Bei Fehlen einer linearen Beziehung wird er $0,0$, bei einem maximal gegensätzlichen Zusammenhang dagegen erhält man $-1,0$. Dieser in den Sozialwissenschaften häufig verwendete statistische Kennwert wird auch als „Produkt-Moment-Korrelation“ bezeichnet und lässt sich für zwei Standardvariablen sehr einfach, nämlich als mittleres Produkt der z-Werte, bestimmen (gebräuchliche Formeln zur Berechnung einer **Korrelation** findet man in statistischen Lehrbüchern oder im Internet):

$$r = \frac{1}{n} \cdot \sum (z_y \cdot z_x)$$

(Σ = Summenzeichen, n = Anzahl der Fälle, z_y und z_x = Standardwerte der Variablen Y und X)

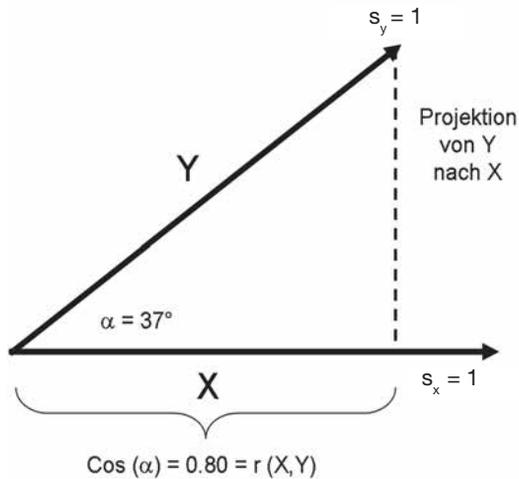
Eine praxisrelevante Nutzenanwendung dieser Statistik besteht in ihrer Vorhersagefunktion für die Ausprägungen einer Variablen (Y), wenn die Werte einer anderen, mit ihr (linear) korrelierenden Variablen (X) bekannt sind:

$$z_y = r \cdot z_x$$

Da mittels dieser (linearen) Funktion die Werte einer Variablen auf jene einer anderen Variablen zurückgeführt werden können (Abb. 3.8), nennt man diese Bezugsherstellung **Regression** und das statistische Verfahren **Regressionsrechnung**. Mittels der Korrelation lässt sich somit der vermutete Einfluss einer Variablen auf eine andere Variable abschätzen. Das Ausmaß des statistischen Effektes einer Variablen auf eine oder mehrere andere Variablen wird als **Effektstärke** bezeichnet. So etwa kann man aus dem Quadrat des Korrelationswertes die Stärke des vermuteten Einflusses einer Variablen auf eine andere abschätzen (z.B. $r = 0,5$, $r^2 = 0,25$, d.h. 25 % Prädiktion), wenn es sich um eine bidirektionale (notwendige und hinreichende) Beziehung handelt (s. Abb. 3.3), was bei psychologischen Effekten eher selten der Fall ist. Neben dem Korrelationskoeffizienten existieren noch weitere Kennwerte für Effektstärken (s. Bortz & Döring, 1995; Westermann, 2000).

lat. *regredere*: zurückgehen, zurückführen

Abb 3.8



Wenn zwei Variablen (X, Y) durch zwei Einheitsvektoren symbolisiert werden (d.h. als Standardvariablen mit Standardabweichung von $s = 1$), und die Variablen miteinander im Ausmaß von $r = 0,80$ korrelieren, dann kann diese Relation durch einen Winkel von 37° zwischen den Vektoren \vec{Y} und \vec{X} im Variablenraum dargestellt werden: $r = 0,80 = \cos(37^\circ)$. Der Wert r entspricht somit der Abbildung einer Variablen auf eine andere.

Eine Besonderheit der geometrischen Betrachtungsweise von Variablen besteht darin, dass das Ausmaß ihrer linearen Beziehung (Korrelation) durch den Winkel ihrer Vektordarstellungen im Variablen- bzw. Merkmalsraum dargestellt werden kann (Andres, 1996; Abb. 3.8). Da jeder Vektor eine variablen-spezifische Eigenschaft symbolisiert und gleichgerichtete Bündelungen von Vektoren somit auf Eigenschaftsüberlappungen der entsprechenden Variablen hinweisen, können für solche Variablencluster gewissermaßen „Schwerpunktvektoren“ berechnet werden, die man **Faktoren** nennt und die als oberbegriffliche Beschreibungen der durch die Variablen symbolisierten Eigenschaften aufzufassen sind (Abb. 3.9).

Mittels solcher faktorieller Beschreibungen kann man nicht nur komplexe Variablensysteme auf ihre „Hauptkomponenten“ reduzieren, sondern auch den korrelativen Zusammenhang zwischen verschiedenen Gruppen von Variablen (mit ähnlicher Eigenschaftsbedeutung) bestimmen. (Statistische Verfahren, die auf diesem Prinzip basieren, sind etwa die „Faktorenanalyse“, die „Multivaria-

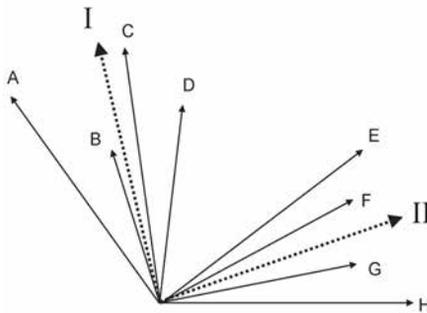


Abb 3.9

Das Prinzip der „Faktorenanalyse“: Wenn zwischen je zwei dieser acht Variablen der Korrelationskoeffizient berechnet wird und die Variablen in den entsprechenden Winkeln zueinander grafisch dargestellt werden, können Bündel davon durch sogenannte Faktoren (I, II) charakterisiert werden. Die vorliegenden acht Variablen lassen sich relativ gut in nur zwei Dimensionen darstellen, wobei die Länge der Variablenvektoren das Ausmaß ihrer Charakterisierbarkeit durch die beiden senkrecht zueinander stehenden Faktoren widerspiegelt. Im Beispiel könnten die vier Variablen A, B, C und D etwa die Eigenschaften schön, vielfältig, harmonisch und heiter von architektonischen Objekten symbolisieren und aufgrund ihrer vektoriellen Bündelung einen Faktor (I) beschreiben, den man ästhetischer Eindruck nennen könnte.

te Varianzanalyse“, die „Kanonische Korrelation“ oder die „Diskriminanzanalyse“.)

Eine Erweiterung dieser Verfahren ist die sogenannte „topologische Datenanalyse“ (Wasserman, 2018; Morris, 2015), bei der Daten an empirische Formen oder Strukturen angepasst werden (z.B. Protein-Strukturen, Kommunikationsnetze).

Inferenzstatistik – schließende und prüfende Statistik

3.6.2

Wie mehrfach erwähnt, müssen in der Psychologie Schlussfolgerungen über die allgemeine Gültigkeit von Gesetzen auf Basis von *Stichproben* gezogen werden. Dies geschieht zumeist unter Verwendung der **Wahrscheinlichkeitstheorie**, mittels derer man zu bestimmen versucht, ob die in den Daten festgestellten Variablenrelationen nur zufällig oder doch durch Einwirkung eines Gesetzes zustande gekommen sind.

Vereinfacht, aber sehr prägnant kann das Bestreben empirischer Sozialforschung anhand des mathematischen **Bayes-Theorems** illustriert werden:

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

lat. a posteriori: von dem, was nachher kommt

lat. a priori: von vornherein, ohne Einbezug von Erfahrungen

In empirischen Wissenschaften geht es um die Einschätzung der Wahrscheinlichkeit $p(H|D)$ für die Gültigkeit einer Hypothese (H) unter der Bedingung, dass hypothesenbestätigende (oder widerlegende) empirische Daten (D) berücksichtigt werden. Die „Aposteriori-Wahrscheinlichkeit“ $p(H|D)$ für eine Hypothese (d.h. nach Einbezug der Daten) nimmt zu, wenn die „Apriori-Wahrscheinlichkeit“ für die Hypothese $p(H)$ größer wird und/oder wenn die Wahrscheinlichkeit $p(D|H)$ für das Auftreten hypothesenbestätigender Daten ebenfalls zunimmt. Sie nimmt hingegen ab, wenn die hypothesenrelevanten Daten auch unabhängig von der Hypothese häufiger auftreten, das heißt, wenn $p(D)$ größer wird.

Die Plausibilität dieses Ansatzes kann am Beispiel einer medizinischen Diagnose über das Vorliegen einer Covid-19-Infektion illustriert werden: Die Annahme, dass eine Person an Covid-19 (C) erkrankt ist, wenn sie Fieber hat ($p(C|F)$), stimmt umso eher, (1) je größer $p(C)$ ist, das heißt, je mehr Personen bereits an Covid-19 erkrankt sind (z.B. bei einer Epidemie), (2) je größer $p(F|C)$, die Wahrscheinlichkeit von Fieber bei dieser Viruserkrankung, ist und (3) je kleiner $p(F)$ ist, nämlich die Erwartung des Auftretens von Fieber im Allgemeinen (s. auch 8.5.3; Tschirk, 2019).

Merksatz

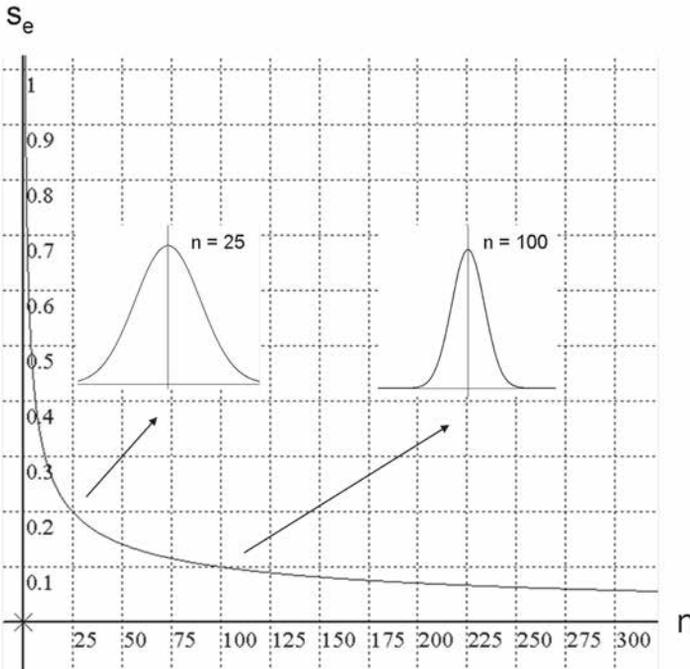
Die möglichst stabile Kennzeichnung von Personen oder Personengruppen hinsichtlich wichtiger Eigenschaften, Einstellungen oder Handlungsweisen („Punktschätzungen“) ist eine zentrale sozialwissenschaftliche Zielsetzung.

Eine zentrale sozialwissenschaftliche Zielsetzung besteht in der möglichst stabilen Kennzeichnung von Personen oder Personengruppen hinsichtlich wichtiger Eigenschaften, Einstellungen oder Handlungsweisen („Punktschätzungen“). Da solche Kennwerte immer fehlerbehaftet sind, wird mittels statistischer Techniken ein **Vertrauensintervall** bzw. **Konfidenzintervall** für sie bestimmt, innerhalb dessen mit

95%iger (99%iger) Wahrscheinlichkeit der „wahre“ Kennwert vermutet wird.

Es ist leicht einzusehen, dass der **Schätzfehler** für einen statistischen Kennwert mit zunehmender Größe der Stichprobe immer kleiner wird und schließlich gegen Null geht, wenn alle möglichen Fälle in die Berechnung einbezogen sind (Abb. 3.10).

Abb 3.10



Der Schätzfehler (s_e) für die Bestimmung des Mittelwertes einer Population von Fällen aufgrund einer Stichprobe ist eine Funktion der Stichprobenstreuung (s) und des Stichprobenumfanges (n): $s_e = s/\sqrt{n}$. Je mehr Fälle für eine Schätzung zur Verfügung stehen, desto genauer wird die Vorhersage. Wenn etwa geschätzt werden sollte, wie viel Zeit Arbeiterinnen und Arbeiter durchschnittlich für einen bestimmten Arbeitsgang in einem Produktionsprozess benötigen, dann wird die Schätzung des Mittelwertes anhand einer Stichprobe von 100 Personen eine nur halb so große Fehlerstreuung aufweisen (in Einheiten der Standardabweichung) wie jene auf Basis einer Stichprobe von 25 Personen.

Die mathematisch begründeten Methoden der **Inferenzstatistik** sollen also eine Einschätzung erlauben, ob überhaupt und in wel-

lat. inferre: hineinragen

chem Ausmaß statistische Resultate von Stichproben auf die jeweilige Population übertragbar sind.

Wenn die Wahrscheinlichkeit dafür, dass bestimmte Variablenrelationen zufällig zu erklären sind, einen vereinbarten Wert unterschreitet (z.B. $p = 0,05$, $p = 0,01$ oder $p = 0,001$), dann spricht man von **statistischer Signifikanz** des Ergebnisses. Bortz und Döring (1995, 27) definieren statistische Signifikanz als ein „per Konvention festgelegtes Entscheidungskriterium für die vorläufige Annahme von statistischen Populationshypothesen“. Wenn also ein statistisches Ergebnis nur mehr zu 5 % (oder weniger) durch Zufallsprozesse erklärt werden kann, wird es als statistisch signifikant angesehen („überzufällig“ oder „unterzufällig“). Die restliche, für eine Zufallserklärung verbleibende Unsicherheit von 5 % (oder weniger) nennt man **Irrtumswahrscheinlichkeit** („Fehler 1. Art“, „Alpha-Fehler“), die dazugehörige den Zufallsprozess charakterisierende Annahme (über die Datenverteilung) heißt **Nullhypothese**.

Da die praktische Bedeutsamkeit eines signifikanten Ergebnisses aber auch von dessen *Effektstärke* abhängt, müssen abgesehen von der Nullhypothese auch **Alternativhypothesen** statistisch getestet werden. Das Ausmaß, in dem die Datenverteilungen mit den Vorhersagen einer Alternativhypothese übereinstimmen, wird als **Teststärke** (engl. power) bezeichnet. Um diese berechnen zu können, ist es nötig, die jeweilige Alternativhypothese zu spezifizieren, indem man die erwartete *Effektgröße* präzisiert, d.h. schätzt, wie stark die jeweilige unabhängige Variable auf die abhängige Variable einwirken dürfte. Der Vorteil einer solchen Vorgangsweise besteht vor allem darin, dass man nicht nur vage auf „Über- oder Unterzufälligkeit“ von statistischen Ergebnissen schließt, sondern sogar die Wahrscheinlichkeit bestimmen kann, mit der die Daten für die Alternativhypothese sprechen.

Merksatz

Inferenzstatistische Verfahren zielen darauf ab, den Grad der Allgemeingültigkeit von Gesetzmäßigkeiten zu prüfen, die auf Basis von Stichproben gewonnen werden.

Forschungsmethoden der Psychologie

| 3.7

Laborexperiment

| 3.7.1

Mittels eines Experiments ist es möglich, hypothetische Wirkfaktoren gezielt zu manipulieren, um ihre Auswirkungen unter verschiedenen Bedingungen zu analysieren. Experimente werden bevorzugt zur Prüfung von Kausalhypothesen eingesetzt (Stapf, 1987). Im Experiment wird eine künstliche Realität konstruiert, um die vermuteten Einflussfaktoren in ihrer Wirksamkeit unter Abschirmung von möglichen Störeinflüssen zu untersuchen.

Häufig wird in psychologischen Experimenten der (den) **Experimentalgruppe(n)** (Versuchsbedingungen) eine **Kontrollgruppe** (Kontrollbedingung) gegenübergestellt. Den Fällen der Experimentalgruppen sind solche Ausprägungen der unabhängigen Variablen (Ursachenvariablen) zugeordnet, von denen ein Effekt auf die abhängigen Variablen (Wirkungsvariablen) erwartet wird, während den Fällen der Kontrollgruppe Ausprägungen der unabhängigen Variablen zugeteilt sind, denen kein systematischer Effekt zugeschrieben wird. Diese Gruppe dient somit nur dazu, Veränderungen zu erfassen, die entweder auf natürliche Weise auftreten (Zeiteffekte, Gewöhnungsprozesse etc.) oder durch die experimentellen Umstände selbst zustande kommen, nämlich durch die künstliche Situation oder den Eindruck, beobachtet zu werden.

Die künstliche Realität des Experiments ist einerseits ein Vorteil, weil durch die Beseitigung von Störeinflüssen der Zusammenhang zwischen unabhängigen und abhängigen Variablen klarer erkannt werden kann (hohe „interne Validität“), andererseits aber auch ein Nachteil, weil die Ergebnisse nur mit Vorsicht auf den Alltag übertragbar sind (geringe „externe“ bzw. „ökologische Validität“).

Ein wesentliches Merkmal psychologischer Experimente ist die **Randomisierung**. Durch die Randomisierung sollen sich Störeffekte ausmitteln, die eventuell durch unausgewogene Stichproben zustande kommen. In der zuvor erwähnten Studie über die Wirkung des Alkohols auf das Fahrverhalten (Box 3.1) würden zum Beispiel die sich meldenden Versuchspersonen per Zufall den Gruppen mit unterschiedlicher Alkoholaufnahme zugewiesen werden.

Die **Störeffekte** in psychologischen Experimenten haben im Wesentlichen drei verschiedene Quellen (Gniech, 1976):

Randomisierung meint die zufällige Zuordnung von Personen (oder Gruppen) zu den jeweiligen Ausprägungen der unabhängigen Variablen.

1. **Versuchssituation:** Der sogenannte „Aufforderungscharakter“ eines experimentellen Umfelds, nämlich die Art der Information über den Zweck der Untersuchung, die Rahmenbedingungen, die Art der Instruktion, die gestellten Fragen und Ähnliches hinterlassen bei den Versuchspersonen Eindrücke, die ihr experimentell induziertes Verhalten beeinflussen können.
2. **Versuchspersonen:** Eine unüberlegte, nicht randomisierte Auswahl der Stichprobe kann Verfälschungen in den Ergebnissen bewirken. Ein nicht zu unterschätzendes Problem bei der Interpretation von Untersuchungsergebnissen ist zum Beispiel die oft notwendige Beschränkung der Teilnahme auf Freiwilligkeit und das Ausscheiden von Teilnehmenden aus dem Experiment („drop out“), wodurch natürlich die erwünschte Zufallsauswahl einer Stichprobe beeinträchtigt ist. Personen, die sich freiwillig für ein Experiment melden, sind im Allgemeinen besser gebildet, haben einen höheren gesellschaftlichen Status, sind stärker sozial orientiert und haben ein stärkeres Bedürfnis nach Anerkennung (Rosenthal & Rosnow, 1975; zit. nach Gniech, 1976). Natürlich wirken sich auch Einstellungsunterschiede der Teilnehmerinnen und Teilnehmer gegenüber der Untersuchung aus, je nachdem, ob es sich um kooperierende, sabotierende oder neutrale Versuchspersonen handelt.
3. **Versuchsleitung:** Vonseiten der Forscherinnen und Forscher sollten **beobachterabhängige Urteilsverzerrungen** (engl.: observer bias) beachtet werden, die durch persönliche Motive und Erwartungen entstehen. Besonders störend sind unbewusste Einflussnahmen (z.B. über nonverbale Kommunikation) im Sinne eigener theoretischer Vorstellungen („Erwartungseffekte“, „Rückschaufehler“, *Selbsterfüllende Prophezeiung*). In Experimenten mit Volksschulkindern (Box 3.4) konnte etwa nachgewiesen werden, dass Lehrpersonen gegenüber fremden Kindern, die ihnen aufgrund von Testresultaten als angeblich begabt ausgewiesen wurden (als „Spätentwickler“), sich sympathischer, förderlicher und entgegenkommender verhielten, sodass sie mit ihrem Verhalten de facto dazu beitrugen, die Fähigkeiten der Kinder zu steigern (Rosenthal & Jacobson, 1968).

Dass Erwartungshaltungen, zum Beispiel hinsichtlich der Wirksamkeit eines Medikamentes, beachtliche Auswirkungen haben können, ist seit Langem aus der Medizin unter der Bezeichnung **Placebo-Effekt** bekannt. Darunter versteht man die positive Wirkung

lat. placebo: „Ich werde gefallen.“

Selbsterfüllende Prophezeiung

| Box 3.4

Die Selbsterfüllende Prophezeiung wird auch Pygmalion-Effekt genannt, nach dem Bildhauer der griechischen Mythologie, der die Statue einer perfekten Frau schuf („Galatea“) und sie durch seinen festen Glauben und seine Sehnsucht nach ihr zum Leben erweckte (Göttin Aphrodite soll allerdings mitgeholfen haben).

In einem Experiment der Sechzigerjahre waren 18 Klassen einer Volksschule einbezogen. Bei allen Schülerinnen und Schülern wurde ein nonverbaler Intelligenztest durchgeführt, den man als Indikator für eine zu erwartende intellektuelle Entwicklung der Kinder in den nächsten acht Monaten ausgab. Aus allen Klassen wurden 20 % der Kinder zufällig (!) ausgewählt, die den Lehrpersonen mit dem Hinweis genannt wurden, dass von diesen Kindern aufgrund des durchgeführten Tests in der nächsten Zeit ein intellektueller Fortschritt zu erwarten sei. Nach acht Monaten zeigten die mit dem positiven Vorurteil bedachten Kinder im Intelligenztest deutliche Verbesserungen! Der gleiche Effekt konnte auch in Tierexperimenten nachgewiesen werden (Rosenthal, 2002).

auf Befinden oder Gesundheit ausgehend von medizinisch unwirksamen Substanzen – sogenannten „Placebos“ (z.B. Milchezucker, Stärke, Salzlösungen) –, allein durch Herstellung einer Erwartung von Wirksamkeit.

Das Gegenteil vom Placebo-Effekt ist der **Nocebo-Effekt**, nämlich die durch Erwartung hervorgerufene negative Auswirkung eines eigentlich unwirksamen Medikaments oder einer Behandlung.

Um die genannten Artefakte in Experimenten zu reduzieren, werden **Doppel-blind-Verfahren** eingesetzt, bei denen weder die Versuchspersonen noch die unmittelbar das Experiment betreuenden Forscherinnen und Forscher über die Art der experimentell gesetzten Einwirkungen Bescheid wissen dürfen. Da natürlich auch in Blindstudien die Probandinnen und Probanden über die Intention einer Studie Vermutungen entwickeln, müssen in der psychologischen Forschung manchmal auch **Täuschungen** eingesetzt werden. Selbstverständlich sind diese nach Ende des Experiments aufzuklären.

Experimente sowie andere Forschungsdesigns können sowohl als **Querschnittuntersuchung** (engl.: cross sectional study) als auch als **Längsschnittuntersuchung** (engl.: longitudinal study) durchgeführt werden. Bei der häufig eingesetzten Querschnittstudie werden an einzelnen Fällen (Personen, Gruppen, Situationen etc.) die interessierenden Variablen nur einmalig erhoben, sodass strukturelle Gesetze von Variablen analysiert werden können, während bei einer Längsschnittstudie zwei oder mehr Datenerhebungen zu den gleichen Variablen stattfinden und somit auch deren zeitliche Dynamik erfassbar ist. Ein großer Vorteil der Längsschnittmethode liegt auch darin, dass intraindividuelle Veränderungen beobachtet werden

können und Verfälschungen durch unausgewogene Stichproben, wie sie bei der interindividuellen Querschnittmethode vorkommen, reduziert sind (Daumenlang, 1987). Nachteilig ist hingegen über einen längeren Zeitraum der Schwund an Versuchspersonen und die Problematik der mehrmaligen Anwendung gleichartiger Testverfahren (Gefahr von „Serieffekten“).

Merksatz

In einem Experiment werden unter abgeschirmten Bedingungen die Effekte (abhängige Variablen) systematisch variiert. Wirkfaktoren (unabhängige Variablen) registriert, wobei durch eine zufällige Zuteilung der Fälle zu den Bedingungen der Wirkfaktoren etwaige Verfälschungen der Ergebnisse minimiert werden sollen.

3.7.2 | Quasiexperiment

Merksatz

Ein Quasiexperiment gleicht vom Aufbau her einem Experiment – mit dem Unterschied, dass die Fallzuordnung zu den Bedingungen nicht zufällig erfolgt.

Diese Untersuchungstechnik gleicht vom Design her dem Experiment, nur verzichtet man auf eine zufällige Zuordnung der Versuchspersonen zu den Versuchs- bzw. Kontrollbedingungen und nimmt das Risiko von Stichprobeneffekten in Kauf. In manchen Forschungsbereichen ist eine zufällige Zuteilung zu den verschiedenen

Bedingungen entweder nicht realisierbar oder ethisch nicht zu rechtfertigen; so etwa die zufällige Zuordnung von Schülerinnen und Schülern zu Schultypen, von Mitarbeiterinnen und Mitarbeitern zu Betrieben oder von Patientinnen und Patienten zu Behandlungsverfahren. Um die aus dem Verzicht einer Randomisierung resultierenden **Artefakte** zu kompensieren, werden in solchen

Artefakt: In Psychologie und Nachrichtentechnik steht dieser Ausdruck für verfälschte Ergebnisse.

Untersuchungen einerseits größere Probandengruppen angestrebt und andererseits zusätzliche Personenmerkmale erhoben, denen ein direkter oder indirekter Einfluss auf die abhängigen Variablen zugeschrieben werden kann. Zu diesen gehören die **soziodemografischen Merkmale** (Geschlecht, Alter, Schulbildung, Beruf ...), aber auch andere individuelle Charakteristika, die aufgrund ihrer Ungleichverteilung in den Bedingungen der unabhängigen Variablen zu systematischen Verfälschungen von Ergebnissen führen könnten. Mittels statistischer Korrekturverfahren lassen sich einige solcher Verfälschungen kompensieren bzw. aus den Ergebnissen herausrechnen („auspartialisieren“).

Feldforschung

| 3.7.3

Im Gegensatz zum Experiment versucht man in der Feldforschung, Phänomene unter möglichst natürlichen Bedingungen zu beobachten und zu erklären. Dem Vorteil der Natürlichkeit steht hier der Nachteil gegenüber, dass Störvariablen weniger gut kontrolliert werden können. Da Forschungsphänomene „im Feld“ wesentlich komplexer in Erscheinung treten als im Labor, kommt bei der Feldforschung der Entwicklung von genauen und effizienten Beschreibungsmethoden sowie der Ausarbeitung von Verhaltensregeln zur optimalen Datengewinnung besondere Bedeutung zu (s. Flick et al., 1995).

Merksatz

Methoden der Feldforschung bezwecken eine Untersuchung von Phänomenen unter natürlichen Rahmenbedingungen bzw. unter minimierter versuchsbedingter Beeinflussung.

Sogenannte **Fallstudien** („single case studies“) sind häufig erste Erfahrungsquellen und als solche nur Anregungen für weitere Forschungstätigkeiten. Obwohl Forschungsphänomene durch Fallstudien hervorragend konkretisiert und plastisch vorstellbar gemacht werden können, mangelt es ihren Ergebnissen logischerweise an Verallgemeinerbarkeit.

Einen Katalog möglicher Verhaltensweisen in natürlichen Umweltbedingungen nennt man in der Verhaltensforschung **Ethogramm**, innerhalb dessen ein „behavior mapping“ charakterisiert, wer was wo tut (Hellbrück & Fischer, 1999).

Non-reaktive Verfahren bezwecken eine Analyse psychologischer Problemstellungen, ohne dass die untersuchten Personen bemer-

ken, dass sie untersucht werden, was insbesondere bei **Inhaltsanalysen** von schriftlichen Dokumenten (z.B. Tagebüchern, Archiven), bei Auszählungen von Unfall-, Krankheits- und Fehlzeitstatistiken, Verkaufsstatistiken oder Abnützungen von Bodenbelägen, Pfaden oder Gebrauchsgegenständen („Spurenanalyse“) gut gelingt.

In der Feldforschung werden häufig, aber nicht ausschließlich, qualitative Methoden verwendet, weil diese flexibler auf die Eigenarten von Personen oder von Situationen anzupassen sind.

3.7.4 | Test und Rating

Merksatz

Eine Standardisierung besteht aus Maßnahmen, die eine Vergleichbarkeit von verschiedenen Personen, Objekten, Situationen oder von Variablenwerten ermöglichen.

Ein Test ist ein wissenschaftlich normiertes und standardisiertes Prüfverfahren, welches stabile Eigenschaften eines komplexen Systems (Person, Gegenstand, Organisation) ermitteln soll. Unter **Standardisierung** versteht man Maßnahmen, aufgrund derer Situationen, Aktionen oder Objekte unter Bezugnahme auf Normen

oder Regeln miteinander verglichen werden können. So etwa müssen in Experimenten Instruktionen und Rahmenbedingungen der Durchführung für alle Versuchspersonen standardisiert, d.h. als maximal ähnlich aufgefasst werden; Gleiches gilt für die Diagnostik, wo Tests verschiedenen Kandidatinnen oder Kandidaten vorgegeben werden. Bei standardisierten Fragebögen müssen die Fragen immer den gleichen Wortlaut haben, auch die möglichen Antworten sind fix vorgegeben (z.B. in Form von Antwortalternativen). Bei statistischen Auswertungen gelten Variablen dann als standardisiert, wenn ihre Werte als Differenz zu ihrem *Mittelwert* und in Einheiten ihrer *Streuung* dargestellt werden

(s. 3.6.1), wodurch auch inhaltlich sehr verschiedenartige Variablen miteinander in Relation gesetzt werden können. Bei Leistungs-, Intelligenz- oder Persönlichkeitstests bedeutet eine Standardisierung, dass die Ergebnisse der Probandinnen und Probanden auf die Verteilungen von sogenannten **Referenz-** oder **Normstichproben** bezogen sind.

Merksatz

Eine Skala soll Ausprägungen einer spezifischen Eigenschaft eines empirischen Sachverhaltes exakt (anhand von Zahlen) charakterisieren.

Da die in einem Test zu erfassenden *Konstrukte* aus Teilaspekten bzw. verschiedenen Inhaltskomponenten bestehen, setzen sich Tests aus entsprechend vielen **Subtests** bzw. **Skalen** zusammen, die jeweils ein homogenes Merkmal feststellen oder „messen“ sollen. Eine Skala ordnet somit empirischen Objekten (z.B. Personen) Zahlen zu, ähnlich wie dies bei der Längenmessung physikalischer Objekte anhand einer Meterskala geschieht (Niederée & Narens, 1996). Die Prüfung, ob zur Vermessung eines empirischen Objekts eine quantitative Skala akzeptiert werden kann, erfolgt auf Basis mathematisch-statistischer **Messtheorien** (s. auch 3.6.1, *Skalenqualität*).

engl. scale: Maßstab, Anzeige, Skala; ital. scala: Maßstab, Treppe, Leiter, Skala

Subtests oder Skalen bestehen selbst meist wieder aus zwei, drei oder mehr Elementen, den **Items**. Je nach inhaltlicher Orientierung des Tests oder der Skala können sich die Items aus Leistungsaufgaben, Fragen mit Antwortalternativen oder aus *Skalierungen*, nämlich Einschätzungen von Merkmalen anhand von Zahlen, zusammensetzen. Die Art der Reaktion einer Person auf ein Item wird über (Zahlen-)Symbole kodiert, welche unter Verwendung mathematisch-statistischer Modelle zu quantitativen Werten (z.B. Mittelwert über die einzelnen Items) für die einzelnen Skalen verrechnet werden. Je mehr Items für eine Skala zur Verfügung stehen, d.h., je mehr unabhängige elementare „Messinstrumente“ für eine Eigenschaft vorliegen, desto größer ist im Allgemeinen die Zuverlässigkeit der entsprechenden Skala.

engl. item: Datenelement, Einheit, Einzelheit, Element, Punkt, Nummer

Eine Aufzählung nach Bühner (2010) soll illustrieren, in welchen verschiedenen Bereichen psychologische Tests eingesetzt werden: psychische und psychosomatische Störungen, Befindlichkeitsstörungen, Therapie- und Heilungsverlauf, Familien-, Ehe- und Erziehungsberatung, Berufsberatung und Personalauslese, Verkehrseignung (TÜV), Strafvollzug (Haftentlassung), Entwicklungsstörungen, Schulreife, Leistungsstörungen, Hochschuleignung, Produktbeurteilung, Einstellungs- und Motivationsmessung (Arbeitszufriedenheit, Leistungsmotivation) usw. „Die Auswahl und Interpretation von Test- und Fragebogenergebnissen zählen zu den Routineaufgaben in der späteren Berufspraxis“ von Psychologinnen und Psychologen (Bühner, 2010, 11).

Die Genauigkeit, die Aussagekraft und der Vorhersageerfolg von Testergebnissen hängen von sogenannten **Gütekriterien** der Tests ab. Allgemeine und unverzichtbare Gütekriterien von Datenerhebungsinstrumenten sind Objektivität, Reliabilität sowie Validität.

Die **Objektivität** eines Tests kennzeichnet die Unabhängigkeit seines Ergebnisses von der Person, die den Test durchführt. Sie ist besonders hoch, wenn verschiedene Testanwender zu gleichen Testergebnissen kommen. Dafür ist es allerdings nötig, dass die Anwenderinnen und Anwender fundierte testpsychologische Grundkenntnisse und Fertigkeiten besitzen (s. DIN-Norm 33430 für „Berufsbezogene Eignungsdiagnostik“, Hornke & Winterfeld, 2004; Bühner, 2010). Objektivitätsmängel können sowohl durch fehlende Sorgfalt bei der Testdurchführung als auch durch Unterschiede bei der Auswertung oder Interpretation entstehen.

Reliabilität bedeutet Zuverlässigkeit und Genauigkeit eines Tests und ist gegeben, wenn bei wiederholter Anwendung des Tests bei gleichen Probanden auch weitgehend gleiche Ergebnisse zustande kommen. Hinweise auf die Zuverlässigkeit von Tests bekommt man, indem man (1) einen Test (falls möglich) wiederholt vorgibt und dessen Ergebnisse auf Übereinstimmung prüft („Retest-Methode“), oder indem man (2) sogenannte Paralleltests, nämlich Tests mit gleicher Aussagekraft, entwickelt und deren Übereinstimmung bei ein und derselben Personengruppe kontrolliert („Paralleltest-Methode), oder indem man (3) die Teile eines Tests auf Homogenität, d.h. auf inhaltliche Ähnlichkeit prüft („Konsistenzmethode“).

Die **Validität** (Gültigkeit), das wichtigste Gütekriterium eines Tests, gibt an, wie gut er in der Lage ist, das zu messen, was er zu messen vorgibt (z.B. Intelligenz, Motivation, Persönlichkeitsmerkmale). „Inhaltliche Validität“ oder „Augenscheinvalidität“ besitzt

ein Test dann, wenn es aufgrund der Art der Testung (Fragen, Leistungen usw.) offensichtlich ist, welcher Aspekt sich im Testergebnis hauptsächlich niederschlägt (z.B. Additionstest für Rechenfertigkeit, Bildermerkttest für Vorstellungsfähigkeit). Die empirische Validitätsprüfung eines Tests geschieht hauptsächlich durch Berechnung des statistischen Zusammen-

hanges (*Korrelation*) seiner Werte mit einem plausiblen Kennwert („Kriteriumsvalidität“) oder mit einem anderen Test, der den gleichen Aspekt zu messen vorgibt („Konstruktvalidität“). Beispielsweise könnte bei Schülerinnen und Schülern für einen Test über rechnerische Intelligenz die Mathematiknote als Validitätskrite-

Merksatz

Ein Test ist ein wissenschaftlich begründetes, normiertes und bestimmten Gütekriterien unterworfenen Verfahren mit dem Ziel einer quantitativen Erfassung von Merkmalen.

rium oder ein ebenfalls auf Rechenleistungen bezogener anderer Test als Validitätskonstrukt herangezogen werden.

Zwischen den genannten drei Gütekriterien besteht allerdings eine Implikationsbeziehung: Wenn ein Test nicht objektiv ist, kann er nicht reliabel sein, und wenn er nicht reliabel ist, ist er nicht valide. Wenn nämlich bereits die Datenerhebung stark fehlerbehaftet ist, können bei wiederholten Messungen keine gleichen Resultate auftreten, und wenn Letzteres nicht gesichert ist, kann auch die zu messende empirische Eigenschaft nicht befriedigend von anderen Eigenschaften unterschieden werden.

Insbesondere bei der Konstruktion von Tests werden neben Objektivität, Reliabilität und Validität noch weitere, ebenfalls wichtige Gütekriterien überprüft (s. Kubinger, 2003): **Skalierung** (quantitative Interpretierbarkeit der Testwerte), **Normierung** (Vergleichsmöglichkeit mit Bevölkerungsgruppen), **Fairness** (Chancengleichheit für alle Bevölkerungsgruppen), **Ökonomie** (Minimum an Ressourcenverbrauch), **Zumutbarkeit** (Minimum an zeitlicher, emotionaler und psychischer Belastung der Probandinnen und Probanden) und **Unverfälschbarkeit** (geringe Möglichkeit zur willkürlichen Beeinflussung der Testergebnisse durch die Testpersonen).

Tab 3.1

| | sehr | ziemlich | eher | weder noch | eher | ziemlich | sehr | |
|------------|------|----------|------|------------|------|----------|------|----------|
| SCHÖN | 3 | 2 | 1 | 0 | 1 | 2 | 3 | HÄSSLICH |
| RUHIG | 3 | 2 | 1 | 0 | 1 | 2 | 3 | LAUT |
| FREMD | 3 | 2 | 1 | 0 | 1 | 2 | 3 | VERTRAUT |
| TEUER | 3 | 2 | 1 | 0 | 1 | 2 | 3 | BILLIG |
| DUNKEL | 3 | 2 | 1 | 0 | 1 | 2 | 3 | HELL |
| GUT | 3 | 2 | 1 | 0 | 1 | 2 | 3 | SCHLECHT |
| VIELFÄLTIG | 3 | 2 | 1 | 0 | 1 | 2 | 3 | MONOTON |
| GROSS | 3 | 2 | 1 | 0 | 1 | 2 | 3 | KLEIN |

Das Polaritätsprofil ist eine in der Psychologie sehr verbreitete Methode zur Erfassung einstellungsbezogener oder gefühlsmäßiger Reaktionen auf Objekte, Personen oder Situationen. Dabei wird von den Versuchspersonen eine Reihe von Eigenschaften oder Eigenschaftspaaren (ca. 5–25) hinsichtlich ihres Zutreffens zahlenmäßig eingestuft.

In der Philosophie den Tests sehr ähnlich und ebenfalls sehr verbreitet sind **Ratingverfahren**, mittels derer Eigenschaften von Personen, Objekten oder Situationen (z.B. Wahrnehmungen, Ausdruckswirkungen oder Einstellungsintensitäten) anhand von Zahlenzuordnungen quantitativ eingestuft werden. Ein Beispiel dafür ist das **Polaritätsprofil** („Semantisches Differential“; Tab. 3.1). In anderen Ratings bzw. **Skalierungen** wird etwa der Grad an Zustimmung zu Meinungen in Prozentpunkten, die Bewertung von Objekten oder Aspekten in Schulnoten oder eine Präferenzentscheidung mittels Punktesystem angegeben.

3.7.5 | Beobachtung

Die **Selbst-** und **Fremdbeobachtung** zählt zu den ältesten Forschungsinstrumenten der Psychologie. Die wissenschaftliche Beobachtung unterscheidet sich von jener des Alltags durch ihre Theoriegeleitetheit und Systematik. „Unter Beobachtung versteht man das systematische Erfassen von wahrnehmbaren Verhaltensweisen, Handlungen oder Interaktionen einer Person oder Personengruppe zum Zeitpunkt ihres Auftretens“ (Ebster & Stalzer, 2003, 221). Grundsätzlich sollte die Beobachtung als Mittel der Informationsgewinnung in allen Untersuchungen zumindest begleitend eingesetzt werden, und auch die beschriebenen **Gütekriterien** von Tests sollten eigentlich für alle Datengewinnungsverfahren in der Psychologie gelten. So sind auch Beobachtungen einer Objektivitätsprüfung zu unterziehen, indem die Übereinstimmung verschiedener, unabhängiger Beobachterinnen oder Beobachter festgestellt wird.

Merksatz

Wissenschaftliche Formen der Selbst- und Fremdbeobachtung sind theoriegeleitet, systematisiert und den allgemeinen Gütekriterien der Datenerhebung unterworfen.

Die **teilnehmende Beobachtung** ist ein Verfahren, bei dem die forschende Person selbst am untersuchten Geschehen teilnimmt und von den erforschten Personengruppen wie ihresgleichen behandelt werden möchte. Man erwartet sich dadurch sowohl eine lebensnähere Perspektive des beforschten Phänomens als auch tiefere Einblicke in die jeweilige Problematik. Bei **nichtteilnehmender Beobachtung** ist man als Forscher den Untersuchungspersonen gegenüber distanzierter eingestellt und an objektiven Ergebnissen interessiert. **Verdeckte Be-**

Hinsichtlich der Einschätzung der subjektiven Gefährdung scheint überdies ein sogenannter **Risiko-Optimismus** wirksam zu sein. In einer Befragung von Ruff (1990) schätzten die Probanden ihre eigene Gefährdung durch Krebs, Herzinfarkt, AIDS, Arbeitslosigkeit, Verkehrsunfall, Überfall, Umweltverschmutzung und Atomkrieg jeweils geringer ein als jene der übrigen Bevölkerung. Eine Untersuchung der Risikobeurteilung von Atomkraftwerken (Maderthaler et al., 1978) hat überdies ergeben, dass die dauernde Konfrontation mit einem Gefahrenobjekt (in diesem Fall durch die Nähe des Wohnortes) eine Abwertung des Risikos zur Folge hatte (vgl. dazu auch die *Theorie der kognitiven Dissonanz*, Kap. 10).

Einen anderen Einfluss auf das Urteilsverhalten zeigen Tversky und Köhler (1994) mit ihrer **Support-Theorie** auf, welche besagt, dass die subjektive Wahrscheinlichkeit einer Gefährdung auch wesentlich von der Beschreibung der Ereignisse abhängt: Je detaillierter die Gefahren geschildert oder subjektiv vorgestellt werden, desto höhere Wahrscheinlichkeiten werden ihnen in Summe zugeschrieben. Die Auswirkungen des expliziten Beschreibens von Ereignissen („unpacking“) zeigte etwa folgende Studie: Natürliche und unnatürliche Todesursachen (Vorkommenshäufigkeit: 92 % bzw. 8 %) wurden von den Probanden zunächst auf 58 % bzw. auf 32 % geschätzt (ohne die Summierbarkeit auf 100 % zu beachten). Wenn nun die Todesursachen aufgeschlüsselt zu beurteilen waren, betrug plötzlich die Summe der subjektiven Wahrscheinlichkeiten für die natürlichen Todesursachen 73 % (Einzelschätzungen: Herzkrankheiten: 22 %, Krebserkrankungen: 18 %, andere natürliche Erkrankungen: 33 %) und für die unnatürlichen Todesursachen 53 % (Unfall: 32 %, Mord: 10 %, andere unnatürliche Ursachen: 11 %).

Intelligenz – Geistige Leistungsfähigkeit

| 8.6

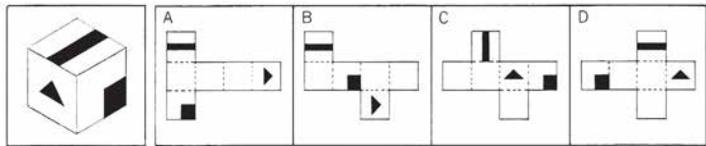
Intelligenz lässt sich definieren als „Fähigkeit, aus Erfahrung zu lernen, Probleme zu lösen und Wissen einzusetzen, um sich an neue Situationen anzupassen“ (Myers, 2005, 460; Spektrum-Lexikon, 2020). Als im Jahre 1905 der französische Bildungs- und Erziehungsminister zur Entwicklung besserer Lehrmethoden für Kinder mit Entwicklungsstörungen aufrief, waren der ursprüngliche Jurist und spätere Psychologe Alfred Binet (1857–1911) und der Arzt Theodore Simon (1873–1961) überzeugt, dass zunächst die Mög-

Merksatz

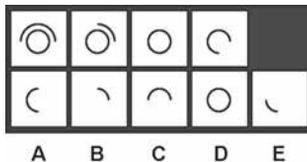
Intelligenz ist ein theoretisches Konstrukt, das die geistige Leistungsfähigkeit einer Person charakterisiert: die erfolgreiche Verarbeitung von Information, der erfolgreiche Einsatz von Lernprozessen und die Nutzung von Wissen zur Lösung von Problemen.

lichkeit einer Selektion der Kinder nach ihrer geistigen Leistungsfähigkeit geschaffen werden müsse. Dies war der Beginn der wissenschaftlich begründeten **Intelligenzdiagnostik** (s. Vanecek, 2003) und der *psychologischen Diagnostik*, welche im „Einsatz festgelegter Testverfahren zur Bewertung von Fähigkeiten, Verhaltensweisen und Persönlichkeitseigenschaften von Personen“ besteht (Zimbardo & Gerrig, 2004, 399). Die eingesetzten Testverfahren müs-

sen heute strengen *Gütekriterien* (s. 3.7.4) genügen, und ihre Weiterentwicklung erfordert daher einen erheblichen zeitlichen und finanziellen Aufwand (Aufgabenbeispiele: Abb. 8.17 und Abb. 8.18).

Abb 8.17

Beispiel einer Intelligenztestaufgabe zur Raumwahrnehmung: Welches der vier Muster (A, B, C oder D) entsteht, wenn der Würfel aufgefoldet wird? (Lösung s. S. 293)

Abb 8.18

Beispiel einer Intelligenzaufgabe zum logischen Schlussfolgern: Welches der Bilder in der unteren Bildreihe stellt eine logische Fortsetzung der obigen Bildfolge dar? (Lösung s. S. 293)

8.6.1 | Intelligenzdiagnostik

Binet (1911) ging mit seiner „Binet-Simon-Intelligenzskala“ ursprünglich davon aus, altersgerechte Intelligenzaufgaben für normal leistungsfähige Kinder zwischen 3 und 15 Jahren zu finden. Die Lösung der Aufgaben einer bestimmten Altersgruppe erlaubte die Zuordnung des entsprechenden „mentalen Niveaus“ bzw. **Intel-**

lignenzalters, welches mit dem Lebensalter in Bezug gesetzt wurde (Herle, 2003). Je größer die Differenz zwischen Intelligenzalter und Lebensalter war, desto mehr wich die Intelligenzleistung vom Durchschnitt der Altersgruppe ab. Da aber bei einem dreijährigen Kind eine Differenz von zwei Jahren mehr bedeutet als bei einem fünfzehnjährigen Jugendlichen, schlug William (alias Wilhelm) Stern (1911/1994) den **Intelligenzquotienten** vor:

$$IQ = \frac{IA}{LA} \cdot 100$$

(Die Multiplikation mit hundert soll ganze Zahlen ermöglichen.)

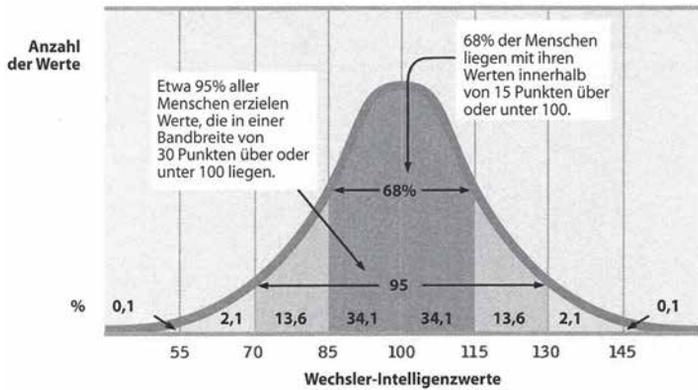
Nachdem weiters festgestellt worden war, dass ab dem frühen Erwachsenenalter bestimmte Intelligenzleistungen wieder abnehmen (analytisches Denken, reasoning, *fluide Intelligenz*, s.u.), musste schließlich auch dieses Intelligenzmaß ersetzt werden durch einen altersrelativierten **Abweichungsquotienten**, wie er bereits von Lewis Terman (1877–1956) in seinem 1916 publizierten „Stanford-Binet-Test“ eingeführt worden war. Dieser bis heute gültige Quotient bezieht die Abweichung der individuellen Testleistung vom Mittelwert auf die Streuung aller Testleistungen der altersentsprechenden Bevölkerung:

$$IQ = 100 + \frac{(\text{Punktwert} - \text{Punktemittelwert})}{\text{Punktstreuung}} \cdot 15$$

Damit wird also eine gruppenbezogene Normierung der Intelligenzwerte vorgenommen. Ein IQ von 100 bedeutet, dass 50 % der altersspezifischen Bevölkerung besser und 50 % schlechter abschneiden (s. Abb. 8.19), und ein IQ-Wert von 70, dass nur etwa 2,5 % der entsprechenden Altersgruppe noch geringere Leistungen aufweisen („Intelligenzminderung“ bzw. „geistige Behinderung“).

Ein interessantes Phänomen ist der **Flynn-Effekt**, der lineare Anstieg in den Intelligenzleistungen seit Beginn der Intelligenzmessung in mindestens 14 Ländern, der bis zu 25 IQ-Punkte von einer Generation zur anderen (30 Jahre) beträgt. Dieser Trend zeigt sich besonders bei kulturunabhängigen Intelligenzleistungen (z.B. *fluid intelligence*) und wird teilweise mit der Verbesserung der Ernährung, der Hebung der Schulbildung, mit zunehmender Umweltkomplexität und Urbanisierung sowie mit ansteigender Testvertrautheit erklärt (Flynn, 1987; Pietschnig, Tran & Voracek, 2013;

Abb 8.19



Da es bei den meisten Leistungs- und Eignungstests keine absoluten, d.h. allgemein gültigen Leistungsniveaus gibt, werden die individuellen Ergebnisse mit jenen anderer Personen verglichen. Dafür wird eine große Stichprobe aus der Gesamtbevölkerung getestet und deren (Normal-)Verteilung zur Bestimmung von Standard- oder Normwerten verwendet (Normierung; s. 3.7.4). Die Abweichung der individuell erreichten Testpunktzahl vom Mittelwert der Bezugsgruppe dividiert durch die Standardabweichung der Verteilung ergibt den Standardwert einer Person (s. 3.6.1) und damit ihre relative Position auf dem allgemeinen Leistungskontinuum. Der Mittelwert aller Intelligenztestwerte ist mit 100 festgelegt und die Streuung mit 15 (oder 10). Aufgrund der angenommenen Normalverteilung der Werte können Prozentsätze für die Intelligenzbereiche bestimmt werden.

Pietschnig et al., 2019). In neueren Veröffentlichungen wird auch ein inverser Flynn-Effekt diskutiert (Dutton et al., 2016).

Im deutschsprachigen Raum sind folgende beiden Intelligenztests sehr gebräuchlich, die jeweils mehrere Intelligenzformen bzw. Intelligenzdimensionen einbeziehen (Testkatalog-Hogrefe, 2018/19):

- Wechsler Adult Intelligence Scale (WAIS-IV)
- Intelligenz-Struktur-Test 2000 R (IST 2000 R)

Einige andere Tests berücksichtigen nur einzelne Intelligenzdimensionen, wie etwa:

- Raven-Matrizen-Test (Standard Progressive Matrices; SPM)
- Dreidimensionaler Würfelttest (3DW)
- Multifactor Emotional Intelligence Scale

Merksatz

Mittels Intelligenztests wird eine quantitative und auf die Altersgruppe normierte Schätzung der geistigen Leistungsfähigkeit angestrebt. Vor einer Überinterpretation des globalen Intelligenzquotienten in einer Weise, als handle es sich um eine genetisch fixierte, angeborene Eigenschaft, muss dringend gewarnt werden.

Wichtig ist es, im Auge zu behalten, dass geistige Leistungsfähigkeit immer nur für einen bestimmten Zeitpunkt und nur für bestimmte Intelligenzformen erfasst werden kann. Die psychische Verfassung spielt dabei eine große Rolle. Zum Beispiel drückt Angst, Stress und vor allem Depression die Messwerte unweigerlich nach unten (s. *Trainingsweltmeister*). Vor einer Überinterpretation des globalen Intelligenzquotienten in einer Weise, als handle es sich um eine genetisch fixierte, angeborene Eigenschaft kann daher nicht eindringlich genug gewarnt werden. Manche Forscher (z.B. Guthke, 2003) unterscheiden überhaupt zwischen „Intelligenzanlage“ (den genetischen Voraussetzungen), „Intelligenzstatus“ (dem momentan erfassbaren Leistungsniveau) und „Intelligenzpotenz“ (potenziell entwickelbare intellektuelle Leistungsfähigkeit).

Um auch die vielfältigen Bedingungen und Voraussetzungen für intellektuelle Leistungen richtig einschätzen zu können („interaktionistische“ Sicht individueller Begabungen; s. 2.3.2), ist für eine zuverlässige Intelligenzdiagnostik auch die zusätzliche Erhebung von leistungsbezogenen Persönlichkeits- und Einstellungsmerkmalen nötig, wie zum Beispiel emotionale Labilität, Belastbarkeit und Leistungsmotivation.

Das in neuerer Zeit bevorzugte computergestützte Testen („computer adaptive/assisted testing“, CAT) machte es möglich, die Testaufgaben von Leistungstests nicht mehr strikt in der gleichen Sequenz vorzugeben, sondern flexibel und maßgeschneidert (und damit zeitsparender) auf das Fähigkeitsniveau der Probandinnen und Probanden anzupassen, wie dies etwa beim AID 3 („Adaptives Intelligenz Diagnostikum 3“) von Kubinger und Holocher-Ertl (2012) der Fall ist. Voraussetzung dafür ist allerdings eine probabilistische Testentwicklung („Item Response Theory“) mit einer wahr-scheinlichkeitstheoretischen Berechnung von Test- und Personen-kennwerten („Itemparameter“ und „Personenparameter“, s. Roskam, 1996; Fischer & Molenaar, 1995).

Formen der Intelligenz

| 8.6.2

Für diagnostische Zwecke wird fast immer ein **Intelligenzprofil** erstellt, das die für die jeweilige Fragestellung (z.B. Berufseignung) interessierenden Stärken und Schwächen einer Person enthält. Die vollständige Erfassung aller intellektuellen Qualitäten eines Men-

schen ist weder theoretisch noch praktisch möglich. Die bis heute psychologisch analysierten Intelligenzformen sind äußerst vielfältig und reichen von einem durch Charles Spearman (1863–1945) postulierten „Generalfaktor“ bis zum „Intelligenzstrukturmodell“ von Guilford (1897–1988) mit 120 Intelligenzaspekten. Für viele solcher Intelligenzbereiche existieren auch psychometrisch entwickelte Testverfahren mit statistisch begründeten **Intelligenzdimensionen**. Häufig eingesetzte Methoden zur Gewinnung statistisch fundierter „Mess- und Strukturmodelle“ (Erdfelder et al., 1996) sind die linearkombinatorische *Faktorenanalyse* und in neuerer Zeit die erwähnten wahrscheinlichkeitstheoretischen Testmodelle („Item Response Theory“; Fischer & Molenaar, 1995).

Eine bereits in den Dreißigerjahren entwickelte, statistisch fundierte Intelligenzkonzeption beruht auf der Idee sogenannter **Primärfähigkeiten** oder **Primärfaktoren** der Intelligenz („primary mental abilities“), von denen angenommen wird, dass sie in unterschied-

Tab 8.4 | „Primary Mental Abilities“ (Thurstone & Thurstone, 1963)

| Benennung der Fähigkeit | Aufgabenart |
|---|--|
| 1. Verbales Verständnis – Verbal comprehension | Erfassen von Wortbedeutungen (Synonyme) |
| 2. Wortflüssigkeit – Word fluency | Lösen von Anagrammen, Bilden von Reimen |
| 3. Rechenfähigkeit – Number | Erkennen korrekter oder inkorrektter Additionen |
| 4. Räumliches Vorstellungsvermögen – Space | Erkennen von zweidimensionalen gedrehten Figuren |
| 5. Merkfähigkeit – Memory | Auswendiglernen von Paarbildungen (Buchstaben und Ziffern) |
| 6. Wahrnehmungsgeschwindigkeit – Perceptual speed | Wiedererkennen von Figuren, Bildvergleiche |
| 7. Schlussfolgerndes Denken – Reasoning | Fortsetzen von Buchstabenreihen (Regeln erkennen) |

Dieses historische Modell der sogenannten „Primärfaktoren“ der Intelligenz wurde erstmals im Jahre 1934 an der University of Chicago durch statistische Analyse (Korrelations- und Faktorenanalyse) der Daten von 240 Personen anhand von 56 Testaufgaben postuliert.

licher Beteiligung die intellektuelle Leistungsfähigkeit des Menschen charakterisieren (Tab. 8.4). Bis zu 20 und mehr solcher Primärfähigkeiten wurden postuliert (Visualisierungskapazität, Informationsverarbeitungsgeschwindigkeit, Einprägungsfähigkeit, Einfallsreichtum usw.) und – ausgehend von ihren internen *Korrelationen* – zu Fähigkeitsdimensionen höherer Ordnung zusammengefasst („Comprehensive Ability Battery“; Hakstian & Cattell, 1978).

Multiple Intelligenzen

Box 8.3

Ausgehend von der Beobachtung, dass bei Gehirnschäden oft sehr spezifische Ausfälle stattfinden, dass bei geistiger Minderbegabung manchmal außergewöhnliche Spezialfähigkeiten in eng begrenzten Bereichen auftreten (z.B. Musik, Rechnen, Raumvorstellung) und dass in Testungen zahlreiche unkorrelierte Intelligenzdimensionen nachgewiesen werden konnten, geht Howard Gardner (2000) von zumindest acht voneinander unabhängigen Intelligenzformen aus:

- Sprachliche Intelligenz (Sprachverstehen, Schreiben, Reden und Lesen)
- Logisch-mathematische Intelligenz (logisches Schlussfolgern, Gleichungen lösen)
- Visuell-räumliche Intelligenz (Stadtpläne interpretieren, Gegenstände ordnen)
- Musikalische Intelligenz (Musikverständnis, Rhythmusgefühl, Gehör, Musizieren, Komponieren)
- Motorisch-kinästhetische Intelligenz (Geschicklichkeit, Körperkontrolle, Tanzen, Turnen, Ballspiele)
- Interpersonale Intelligenz (Einfühlung, Kommunikation)
- Intrapersonale Intelligenz (Selbstverständnis, Kenntnis der eigenen Stärken und Schwächen)
- Naturalistische Intelligenz (Klassifikation und Verständnis von Strukturen und Vorgängen in der Natur)

Obwohl dieses Konzept plausibel erscheint und in viele pädagogische Programme eingeflossen ist, wird es noch nicht als ausreichend empirisch evaluiert angesehen.

Zwei solcher Hauptfaktoren fanden in der Intelligenzforschung besondere Beachtung: Die **kristalline Intelligenz** („crystalized intelligence“), welche kulturabhängige Fähigkeiten erfasst (verbale und motorische Fertigkeiten, Expertenwissen), und die **fluide Intelligenz** („fluid intelligence“), welche sich auf die weitgehend kulturunabhängige kognitive Grundausstattung der geistigen Leistungsfähigkeit bezieht (z.B. induktives, kombinatorisches und figurales Denken). Die wissenschaftliche Produktivität liegt bei Fächern mit hoher Anforderung in fluider Intelligenz (Mathematik, Physik, Chemie) im Alter zwischen 25 und 40 Jahren (Salthouse, 2012).

Eine Ausweitung dieses Konzepts unabhängiger *Intelligenzdimensionen* stellt die **Theorie der multiplen Intelligenzen** von Gardner, 2000; 2006) dar (Box 8.3), bei der auch Fähigkeitsbereiche einbezogen wurden, die bisher in Intelligenztests kaum als solche Berücksichtigung fanden (z.B. Musikverständnis, Selbstkenntnis, Geschicklichkeit).

Merksatz

Geistige Leistungsfähigkeit manifestiert sich in sehr verschiedenen, voneinander weitgehend unabhängigen Bereichen, für die derzeit nur teilweise Testinstrumente zur Verfügung stehen.

Ein häufig geäußerter Kritikpunkt an klassischen Intelligenztests war ihre relativ geringe Prognoseleistung (ca. 25 %) für schulischen oder beruflichen Erfolg (Neisser et al., 1996) und ihre geringe Korrelation mit komplexen Problemlöseleistungen

(z.B. Dörner, 1989; Salthouse, 2012). Eine besondere Annäherung der Intelligenzmessung an Alltagserfordernisse bezweckt das Konzept der **Erfolgsintelligenz** („Theory of Successful Intelligence“) von Sternberg (1997). Sie wird definiert als

- Fähigkeit, im persönlichen Rahmen innerhalb des soziokulturellen Kontextes Erfolg zu haben,
- als Begabung zur Nutzung eigener Stärken und zur Kompensation eigener Schwächen,
- und erfordert eine Balance zwischen *analytischer* (schulischer, akademischer), *kreativer* und *praktischer* Intelligenz (Abb. 8.20).

Einen ebenfalls lebensnahen, in klassischen Intelligenztests kaum berücksichtigten Aspekt beschreibt der Begriff der **emotionalen Intelligenz** (Goleman, 1996), welcher die Geschicklichkeit charakterisieren soll, einerseits die eigenen Emotionen richtig einzuschätzen, mit ihnen vernünftig umzugehen und sie unmissverständlich mitzuteilen, und andererseits auch die Emotionslage anderer richtig zu interpretieren und erfolgreich zu beeinflussen.

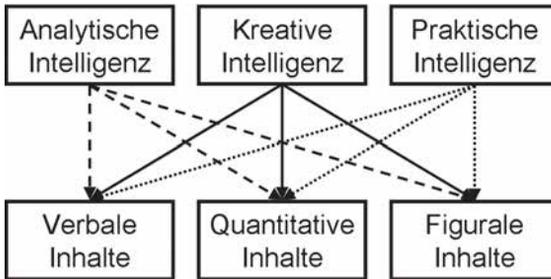


Abb 8.20

In der triarchischen Theorie der Intelligenz („Triarchic Theory of Intelligence“) von Sternberg und Kaufman (1998) wird von drei selbst wieder aus anderen Unterkomponenten bestehenden Intelligenzformen ausgegangen, die sich in je drei Inhalts- bzw. Präsentationsebenen manifestieren. Mittels einer entsprechenden Testbatterie („Sternberg Triarchic Abilities Test“, STAT) konnten diese (neun) Faktoren an einer internationalen Stichprobe von mehr als 3000 Probandinnen und Probanden gut bestätigt werden (Sternberg et al., 2001). Die analytische Intelligenz einer Person (der Hauptaspekt konventioneller Intelligenztests) wird durch kreative und praktische Fähigkeiten („tacit knowledge“) hier ergänzt gesehen, welche nach Sternberg in vielen Lebensbereichen mindestens ebenso wichtig sind (Sternberg, 2002).

Dass bisher kaum Formen emotionaler, sozialer oder praktischer Intelligenz in gängigen Intelligenztests vorkamen, lag wohl weniger an der Unterschätzung ihrer Alltagsrelevanz durch die Psychologinnen und Psychologen, sondern eher an den Schwierigkeiten ihrer Erhebung. Die in den entsprechenden Tests verwendeten Aufgaben (z.B. Erkennen von Stimmungen, Einfühlung in andere, Bewältigung sozialer Konfliktsituationen) lassen oft keine eindeutige Bewertung der Antworten zu und verursachen damit Probleme der *Objektivität*, *Reliabilität* und *Validität* (s. 3.7.4).

Genetische Veranlagung und Umweltfaktoren der Intelligenz

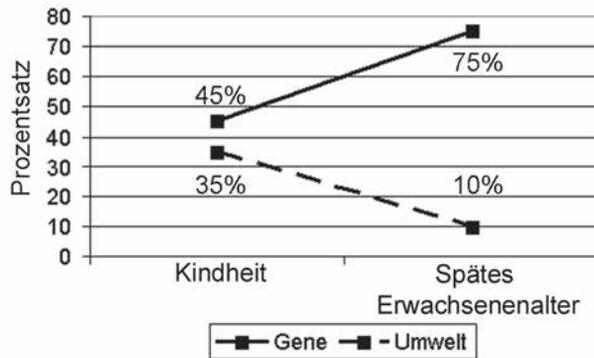
8.6.3

Die **Erblichkeitsschätzung** (Ausmaß des genetischen Einflusses auf ein Persönlichkeitsmerkmal) für Intelligenz ist von einigen verwerflichen und folgenschweren Entgleisungen in der Frühzeit der Intelligenzforschung geprägt. Ohne die Umwelt- und Kulturabhängigkeit von intellektuellen Leistungen zu beachten, wurden zum Beispiel in den USA Immigrantinnen und Immigranten jüdischer,

italienischer und russischer Herkunft, aber auch Latinos und Schwarze aufgrund angeblich niedrigerer Intelligenzwerte diskriminiert und als genetisch „minderwertig“ eingestuft, was im Jahre 1924 zu entsprechenden Einwanderungsbeschränkungen („Immigration Restriction Act“) führte.

Nachdem nicht nur unter nationalsozialistischer Herrschaft, sondern auch später immer wieder ideologisch gespeiste Vermu-

Abb 8.21



Die Einflüsse der Gene und der Umweltfaktoren (Familie, Bildung, Beruf) auf Intelligenzleistungen können mittels statistischer Verfahren geschätzt werden. Vergleiche zwischen Eltern und Adoptivkindern, Eltern und leiblichen Kindern, Geschwistern sowie eineiigen Zwillingen (unähnliches, ähnliches, gleiches Erbgut) in übereinstimmenden oder getrennten Lebensbedingungen erlauben eine Schätzung des Ausmaßes genetischer und umweltbedingter Effekte auf die Variation von Intelligenzmessungen. Nach Neisser und Mitarbeitern (1996, 85) nimmt der Einfluss der Umwelt mit dem Alter ab (35 % → 10 %) und jener der genetischen Disposition zu (45 % → 75%).

tungen über angeblich konstante, ethnisch spezifizierbare Intelligenzausstattungen des Menschen auftauchten, hat eine Arbeitsgruppe von Intelligenzforschern („Board of Scientific Affairs“) der American Psychological Association (APA) in einem wissenschaftlichen Überblicksartikel zu wesentlichen Aspekten der geistigen Leistungsfähigkeit Stellung bezogen (Neisser et al., 1996). Demnach sind die Unterschiede im IQ zwischen US-Amerikanern asiatischer, hispanischer, indianischer und afrikanischer Herkunft minimal

(bis zu 10 Punkten) und gleichen sich im Laufe der kulturellen Eingliederung weitgehend an jene der „weißen“ Bevölkerung an.

Was die allgemeine Frage der genetischen Determiniertheit von Intelligenz betrifft, so wird aus Zwillings- und Geschwisterstudien geschlossen (s. auch Abb. 2.1 in 2.3.2), dass der Anteil an Erbllichkeit für die Leistung in Intelligenztests etwa 50 % beträgt und jener für familiäre oder Umwelteinflüsse etwa 25 % (Rest ist Zufall). Allerdings nimmt mit dem Alter der Einfluss der genetischen Ausstattung zu und jener der Umwelt ab (Abb. 8.21). Dies wird dadurch erklärt, dass mit dem Alter die Selbstständigkeit wächst und somit die Chance, sich jene Umweltbedingungen auszusuchen oder zu schaffen, die der zugrunde liegenden genetischen Ausstattung bzw. den erblich bedingten Neigungen am besten entsprechen („genoty-environment correlation“; Plomin & Spinath, 2004).

Frauen und Männer unterscheiden sich in Gesamt-IQ-Werten im Allgemeinen nicht, sind aber in einigen Intelligenzaufgaben differenzierbar (Abb. 8.22). Insgesamt werden die gefundenen ge-

Merksatz

Der Anteil an Erbllichkeit für die Leistung in Intelligenztests beträgt etwa 50 %, jener für familiäre oder Umwelteinflüsse etwa 25 %. Mit dem Alter nimmt der Einfluss der genetischen Ausstattung zu und jener der Umwelt ab.

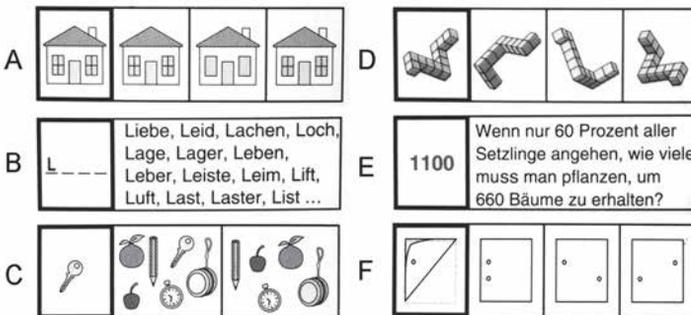


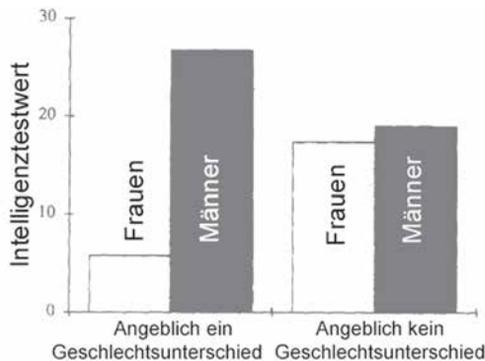
Abb 8.22

Frauen sind durchschnittlich im schnellen Identifizieren von Bildern (A), im Finden von Wörtern (B) und Vergleichen von Objektlisten (C) besser, während Männer im Allgemeinen besser rotierte Figuren identifizieren (D), mathematische Schlussfolgerungen lösen (E) und Strukturveränderungen erkennen (F).

schlechtsspezifischen Intelligenzdiskrepanzen oft übertrieben interpretiert und übermäßig verallgemeinert: Frauen schneiden tendenziell in Verbal- und Kommunikationsleistungen besser ab (Lesen, Rechtschreibung, Ideen- und Wortflüssigkeit, Interpretation der Körpersprache) und weisen weniger Lese- und Sprachstörungen auf, während Männer tendenziell bessere visuell-räumliche und mathematische Leistungen zeigen (Rotation von Strukturen in der Vorstellung, Kartenlesen, mechanische Probleme lösen; Kimura, 1993; Neisser et al., 1996).

Neben den bereits genannten Verfälschungsmöglichkeiten von Leistungstestkennwerten (z.B. durch Kulturabhängigkeit, Trainingseffekte, Umweltstimulation, nicht repräsentative Stichprobenauswahl) sind auch soziale Einflussfaktoren gefunden worden, wie etwa die Hemmung von Leistungen durch ein gesellschaftliches **Stereotyp**. Stereotype sind Kategorisierungen von Personengruppen, wobei bestimmte Konfigurationen von Eigenschaften als typisch angenommen werden. Mehr oder weniger eingestanden

Abb 8.23



In einem Experiment (Steele, 1997) nahmen Schülerinnen und Schüler einer High-school teil, die alle als gleich gut und gleich interessiert in Mathematik angesehen werden konnten. Sie bekamen einen schwierigen Mathematiktest vorgelegt, wobei einer Gruppe gesagt wurde, dass in diesem Test Männer im Allgemeinen besser abschnitten als Frauen, während in einer zweiten Gruppe keine geschlechtsspezifischen Unterschiede suggeriert wurden. Die Ergebnisse zeigten, dass Frauen dann schlechter abschnitten, wenn sie erwarten konnten, dass der Test ihre stereotypiebedingten Schwächen erkennen ließe.

existieren solche Auswirkungen *impliziter Persönlichkeitstheorien* für alle Bevölkerungsgruppen, oft in Form gegensätzlicher Stereotype: Männer – Frauen, Schwarze – Weiße, Fremde – Einheimische etc. Stimmen nun bestimmte Eigenschaften (z.B. Mathematikfähigkeit) bei einer Person (z.B. einer Frau) nicht mit jenen überein, die hinsichtlich des einschlägigen Stereotyps erwartet werden (z.B. „Frauen sind schlecht in Mathematik“), dann resultiert für die Person eine subjektive Bedrohung der Zugehörigkeit zu ihrer Bezugsgruppe. Sie wird infolgedessen entweder ihre Leistungen den stereotypen Erwartungen anpassen oder eine Abschwächung der Identifikation mit dem Leistungsbereich vornehmen (Steele, 1997). Leistungshemmungen, die durch solche **Stereotypbedrohungen** („stereotype threats“) erklärbar sind, konnten bei Frauen in Mathematik und bei Afroamerikanern in Verbaltests eindeutig nachgewiesen werden (Steele, 1997; Steele & Aronson, 1995; Abb. 8.23). Sie treten aber wahrscheinlich in allen Situationen auf, in denen aufgrund sozialer Stereotype negative Leistungserwartungen erzeugt werden.

Zusammenfassung

Der Zweck jeglichen menschlichen Handelns ist im Grunde die kurzfristige oder langfristige Adaptation an Umweltgegebenheiten. Probleme können daher psychologisch als Situationen definiert werden, in denen die gewünschte Überführung von einem Ist- in einen Soll-Zustand mit Aufwand verbunden ist. Die Lösung eines Problems wird durch Informationen über die Problemsituation und ihre Veränderungsmöglichkeiten (über den Problemraum) erleichtert. Theoretisch werden als Grundlage des Problemlösens mentale Repräsentationen (kognitive Abbildungen) des Problemraumes angenommen, um korrektes oder fehlerhaftes Problemlöseverhalten zu erklären.

Wichtige psychische Teilprozesse des Problemlösens sind Schlussfolgerungen, welche entweder induktiv ablaufen, indem aus Einzelerfahrungen auf Regeln geschlossen wird, oder deduktiv, indem aus Prämissen implikative Ableitungen durchgeführt werden. Letztere unterteilen sich wieder in Konditionalschlüsse (Wenn-dann-Formulierungen) und in Kategoriale Schlüsse (Mengenaussagen). Schlussfolgerungen sind fehleranfällig, besonders wenn