

---

## Zusammenfassung

Anhand von Beispielen werden gängige Methoden der Sozialpsychologie möglichst anschaulich eingeführt. Das Kapitel wird grundlegende Begriffe (Theorie, Konstrukt, Variable, Hypothese usw.) erläutern, gängige Untersuchungsstrategien (Experiment, Quasi-Experiment, Umfrageforschung, qualitative Forschung) vorstellen und Risiken für die Qualität der Forschung diskutieren. Schließlich werden ethische Standards psychologischer Forschung aufgezeigt und beispielhaft an einer einzelnen Studie (z. B. BBC-Gefängnis Experiment) diskutiert.

---

## 2.1 Einleitung

Die Sozialpsychologie ist eine Teildisziplin der Psychologie. Wie schon im Eingangskapitel ausgeführt, die Sozialpsychologie untersucht, wie Menschen durch die tatsächliche oder vorgestellte Anwesenheit anderer, in ihrem Denken, Fühlen, Entscheiden und Verhalten beeinflusst werden (s. Allport 1954). Im Gegensatz zu weniger empirisch ausgerichteten Wissenschaftsdisziplinen versucht die Psychologie und damit natürlich auch die Sozialpsychologie, ihre Wege des Wissensgewinns durch Einsatz empirischer Methoden genauer zu beleuchten, kritisierbar zu machen und damit weiter zu entwickeln. In der Sozialpsychologie wird versucht, systematisch und mittels unterschiedlicher empirischer Vorgehensweisen (z. B. experimenteller Methoden) zu untersuchen, wie sich Menschen durch die (vorgestellte) Anwesenheit anderer verändern.

Wenn wir uns überlegen, warum wir wissenschaftliche Methoden in der Sozialpsychologie benötigen, dann können wir Nietzsche zustimmen, wenn er in „Menschlich, allzu menschliches“ schreibt „... die kleinen unscheinbaren, vorsichtigen

Wahrheiten, welche mit strenger Methode gefunden wurden, höher zu schätzen als jene weiten schwebenden umschleiernden Allgemeinheiten, nach denen das Bedürfnis religiöser und künstlerischer Zeitalter greift.“ Wir brauchen strenge Methoden, weil wir damit Wissen erwerben, das allgemein gültig und überprüfbar ist. Die Gültigkeit unseres Wissens sollte also nicht davon abhängen, welchen Ideen wir anhängen, welchen Autoritäten wir uns unterordnen und welcher politischen Agenda wir zuneigen. Um Missverständnissen vorzubeugen: Wissen ist dann nützlich, wenn man damit eigene Ziele erreichen kann. So kann also die subjektive *Nützlichkeit* von Wissen mit unterschiedlichen Zielen variieren, die *Gültigkeit* des Wissens sollte dies jedoch nicht tun.

Peirce (1877) unterscheidet vier verschiedene Quellen unserer Überzeugungen und Meinungen. So kann man den Ansichten und Meinungen von **Autoritäten** folgen. Dies ist einfach und schnell. Es hat zudem den Vorteil, dass man sich selbst nicht genauer mit einem Gegenstand beschäftigen muss (oder kann). Es ist nicht nötig selbst Medizin zu studieren, um eine Erkältung behandeln. Es ist völlig ausreichend einen Arzt zu fragen. Diese Methode hat auch Nachteile, denn wenn sich Experten irren oder gar keine Experten sind, dann schlägt diese Methode fehl. Manchmal ist es schwierig den richtigen Experten zu identifizieren. Sollte man also zum Allgemeinmediziner oder zum Heilpraktiker gehen? Eine zweite Methode des Wissenserwerbs ist die **Alltagserfahrung** und der „gesunde Menschenverstand“. Alltagswissen ist intuitiv überzeugend, denn es basiert auf eigenen Erfahrungen, welche das Gefühl vermitteln, einen bestimmten Umstand bereits erlebt zu haben. Der Nachteil ist dabei, dass unsere Intuition weniger darauf ausgerichtet ist, eine „objektive“ Welt<sup>1</sup> abzubilden, sondern eher darauf, unsere Interessen zu vertreten (siehe Kap. 3 und 5). Aus diesen Gründen ist unser Alltagswissen selten systematisch und objektiv. Eine dritte Art Wissen zu gewinnen ist „**Hartnäckigkeit**“ und „Beharrlichkeit“ (tenacity). Hier bestimmt eher der Wille zum Glauben, ob man einen Inhalt für wahr oder falsch hält. Diese Art des Wissenserwerbs (oder der Mangel desselben) findet sich manchmal in Ideologien, Religionen oder Wissensgebilden, die dadurch stabil werden, dass widersprechende Evidenz einfach wegeklärt wird („Sie können psychoanalytische Konzepte erst dann kritisieren, wenn Sie eine Lehranalyse

---

<sup>1</sup>Wobei wir Davidson (2001) folgend „subjektiv“ ein Urteil aus der Perspektive eines Individuums verstehen, das „intersubjektiv“ durch ein Urteil einer anderen Person bestätigt oder herausgefordert werden kann. „Objektiv“ sind die Urteile, wie sie von vielen (oder allen) Menschen konsensuell abgegeben werden.

hinter sich haben.“ oder „Kritik am Kommunismus ist ungültig, denn die Kritiker sind noch in ihrem bürgerlichen Denken verhaftet“). Der vierte Weg Wissen zu erwerben ist die **wissenschaftliche Methode**. Wesentliche Merkmale der wissenschaftlichen Methode sind, dass sie öffentlich und selbstkorrigierend ist. Öffentlichkeit macht Wissen kritisierbar. Selbstkorrigierend bedeutet hier, dass überkommene Autoritäten, unsere Alltagsintuitionen, unsere bevorzugten Wahrheiten und damit schlicht alles Wissen auf den Prüfstand gestellt, korrigiert und verändert werden kann. Da es einzelnen Forschenden unmöglich ist, alle Verzerrungen durch eigene Wünsche, Intuitionen usw. selbst zu korrigieren, ist Wissenschaft ein kooperatives Unternehmen, an dem viele Forschende teilnehmen und sich gegenseitig korrigieren. Daher ist Wissenschaft gleichzeitig kumulativ. Neue Forschung sollte also immer in Bezug zu vorher Erforschtem und Behaupteten gesetzt werden, und sei es dadurch, dass empirisch widerlegte Vorstellungen (z. B., dass die Körperform die Persönlichkeit bestimmt) verworfen, korrigiert oder ausdifferenziert werden (z. B. kann körperliche Attraktivität die *Wahrnehmung* von Persönlichkeit positiv beeinflussen).

Was sind also die Standards der Wissenschaft? Der erste Standard ist die **Validität**, d. h. unsere Forschung sollte **gültig** sein. So sollten unsere Theorien, Studien, und Messinstrumente das erklären, untersuchen und erfassen, was sie vorgeben zu erklären, untersuchen und erfassen. Da wir bestimmte Phänomene wie beispielsweise Aggression nicht direkt beobachten können sondern immer nur einzelne Fälle von Verhalten, das wir als aggressiv klassifizieren, müssen wir sicherstellen, dass wir wirklich Fälle von Aggression (also intendierte Schädigung Anderer) vor uns haben und nicht etwa verwandtes – aber konzeptuell unterscheidbares – Dominanzverhalten. Der zweite Standard ist die **Reliabilität** (Zuverlässigkeit). Damit ist gemeint, dass unsere Messungen, Effekte und Vorhersagen wiederholbar sein müssen. Wenn eine Studie zwar ein interessantes Ergebnis erbracht hat, dieses sich jedoch nicht replizieren lässt, dann sollten wir das Ergebnis mit Vorsicht interpretieren, denn es scheint nicht reliabel und damit ein Zufallsergebnis zu sein. Nur Ergebnisse, die sich replizieren lassen (manchmal nur unter ganz bestimmten Bedingungen), zählen als wissenschaftliche Erkenntnis. Validität und Reliabilität sind die beiden zentralen Kriterien, denen sich die Wissenschaft unterwirft. Andere Kriterien stehen im Dienste dieser beiden: Forschung ist **kumulativ**. Sie baut auf vorhandenem Wissen auf und addiert neue Erkenntnisse zum bereits bestehenden Wissensfundus. Wissenschaftliche Forschung ist **öffentlich**. Alle neuen Erkenntnisse müssen zumindest anderen Forschenden zugänglich sein und damit kritisierbar werden. Durch kritische Abwägung von Argumenten ist wissenschaftliche Erkenntnis selbstkorrigierend und kann damit immer objektiver werden. Ein weiteres häufig genanntes

Kriterium wissenschaftlicher Erkenntnis ist **Einfachheit** oder **Sparsamkeit** („Ockhams Rasiermesser“). Dieses Kriterium besagt, dass eine Theorie, die mit weniger Annahmen die gleichen Beobachtungen erklären kann als eine alternative Theorie, dieser vorzuziehen ist, denn man sollte ohne Not keine zu komplexen Annahmen beibehalten. Ein Problem dieses Kriteriums ist, dass es häufig von der Vorbildung der Theoretiker abhängt, welche Theorie als die einfachere und sparsamere angesehen wird.

---

## 2.2 Theorien in der Forschung

Warum brauchen wir überhaupt Theorien in der Forschung oder Praxis? Wie könnte unser Wissen ohne Theorien aussehen? Ohne Theorien hätten wir vermutlich eine lange Liste von unzusammenhängenden Fakten, wie es in dem „Handbuch des Nutzlosen Wissens“ sehr schön dokumentiert wird (siehe Kasten). Schon eine Sortierung des Wissens nach Bereichen (Phänomen, vermutete Erklärung, Anwendung) wäre ein erster Schritt in Richtung Theorie. Theorien sortieren und ordnen unser Wissen. Neben der ordnenden Funktion machen Theorien deutlich, wo wir schon Einiges wissen (also welche **Konstrukte** klar sind) und wo noch deutlicher Forschungsbedarf vorliegt. Ferner geben Theorien an, wie die Konstrukte der Theorie vermutlich **zusammenhängen**. Auch das leitet aktive Forschungstätigkeit an. Typische Zusammenhänge, auf die die Forschung zielt, sind kausale Zusammenhänge, die besagen, dass die Veränderung in einer Variablen eine Veränderung in einer anderen nach sich zieht. Theorien erlauben damit Vorhersagen von künftigen Ereignissen und damit auch Handlungsanleitungen für Interventionen. Hier trifft die Aussage nach Kurt Lewin (1952) zu, dass nichts praktischer sei, als eine gute Theorie<sup>2</sup>.

Theorien bestehen also aus Konstrukten und deren Zusammenhängen. Ein Konstrukt ist ein abstrakter Begriff, der viele mögliche Beobachtungen hinsichtlich bestimmter Ähnlichkeiten zusammenfasst. Beispielsweise können unter

---

<sup>2</sup>Das originale Zitat ist für sich sehr interessant und lautet folgendermaßen: „Many psychologists working today in an applied field are keenly aware of the need for close cooperation between theoretical and applied psychology. This can be accomplished in psychology, as it has been accomplished in physics, if the theorist does not look towards applied problems with highbrow aversion or with a fear of social problems, and if the applied psychologist realizes that there is nothing as practical as a good theory“ (Lewin 1952, S. 169).

dem Begriff der „Angst“ so unterschiedliche Dinge zusammengefasst werden wie „Herzrasen“, „Zittern“, „Schwindel“, selbst berichtete Angst und ängstlicher Gesichtsausdruck. Alle diese Indikatoren können Angst anzeigen. Sie können aber auch ohne Angst auftreten. Indikatoren für bestimmte Konstrukte sind selten perfekt, sondern nur mit einer bestimmten (häufig hohen) Wahrscheinlichkeit mit dem Konstrukt verbunden (Brunswick 1955). Konstrukte fassen also einzelne Beobachtungen zusammen. Theorien bestimmen aber auch die Beziehungen zwischen den Konstrukten; diese Beziehungen sind typischerweise kausale Beziehungen (z. B. Frustration macht aggressives Verhalten wahrscheinlicher). Damit lassen sich durch Theorien spezifische Beobachtungen auf allgemeine Gesetzmäßigkeiten zurückführen und damit können, sofern die Theorie stimmt, Vorhersagen über zukünftige Ereignisse getroffen werden.

---

## 2.3 Konstruktvalidität

Um Theorien zu überprüfen müssen zuerst die Konstrukte operationalisiert und gemessen werden. Die Qualität dieser Messungen wird **Konstruktvalidität** genannt. Sie beschreibt die Übereinstimmung der gemessenen oder manipulierten Variablen mit den Konstrukten der Theorie. Für eine gute Messung eines Konstruktes ist es häufig unerlässlich, dass das Konstrukt analysiert wird und seine wesentlichen Bestandteile identifiziert werden. Eine solche Analyse von Konstrukten ist häufig in der Philosophie zu finden, in der Psychologie gehen wir hingegen sorgloser vor und verwenden unser Alltagsverständnis. Das kann unter Umständen zu einiger Verwirrung führen. Wenn beispielsweise unter dem Konzept „Respekt“ von einigen „Mögen“, von anderen „Anerkennung“, „gleiche Rechte“ oder „gute Umgangsformen“ verstanden werden, dann sollten unterschiedliche Studien zu „Respekt“ zu widersprüchlichen Ergebnissen führen (siehe Simon 2007). Zudem können Konstrukte durch unterschiedliche Indikatoren erfasst werden, wie etwa Selbstberichte, Verhaltensbeobachtungen oder physiologische Maße. Um sicher zu stellen, dass diese Indikatoren noch das gleiche Konzept erfassen, sollten unterschiedlichen Studien zu diesem Konzept zu ähnlichen (konvergierenden) Ergebnissen führen.

Insbesondere Selbstberichte können durch soziale Erwünschtheit verzerrt werden. **Soziale Erwünschtheit** („social desirability“) ist die Tendenz so zu antworten, wie der Versuchsleiter es vielleicht erwarten würde, wie es soziale Normen vorschreiben oder wie die eigene Person in vorteilhaftem Licht erscheint. Das ist insbesondere ein Problem, wenn gesellschaftlich oder politisch sensible Themen angesprochen werden, wie beispielsweise Vorurteile oder aggressives Verhalten.

So können Versuchspersonen weniger negativ über andere soziale Gruppen urteilen oder weniger aggressive Tendenzen angeben, weil sie fürchten, dann schlecht da zu stehen. Diese Tendenz zur sozialen Erwünschtheit gefährdet die Konstruktvalidität.

Es gibt unterschiedliche Arten, wie in der Psychologie Konstrukte gemessen werden können. Auf der einen Seite werden durch Selbstbeurteilungsmaße in Fragebögen Meinungen, Einstellungen und Haltungen erfasst. Um beispielsweise Einstellungen zu messen, werden einzelne Items formuliert und zu Skalen zusammengefasst (zu unterschiedlichen Skalierungsmethoden siehe Kapitel Einstellungen).

Neben Fragebögen werden auch immer wieder Interviews eingesetzt, durch die eine genauere Exploration der Ansichten der Untersuchungsteilnehmer möglich ist. Interviews können in freier Form gehalten werden, in denen die Untersuchungsteilnehmer bestimmen, was ihnen wichtig ist. Damit werden theoretische Vorannahmen der Forschenden einerseits minimiert, allerdings verringert sich die Vergleichbarkeit zwischen den Interviews und Systematik der Erfassung bestimmter Themen. Als Kompromiss zwischen systematischer Exploration bestimmter Themen und Vorgaben durch die Untersucherinnen und Untersucher werden Interviews häufig standardisiert oder halb standardisiert (z. B. hinsichtlich der Fragen, des Ablaufplans oder sogar der Antwortkategorien).

Alternativ zu psychologischen Tests können auch Beobachtungen angestellt werden, bei denen Beobachtende entweder heimlich oder als teilnehmender Beobachtende das Verhalten von Menschen in deren alltäglichen Situationen notieren. Wie bei Interviews stellt sich bei der Beobachtung das Problem der Standardisierung der Auswertung, das dadurch gelöst werden kann, dass mehrere Beobachter Verhaltensweisen beobachten oder mehrere Bewerter einen Film oder ein Interview hinsichtlich bestimmter Kategorien auswerten. Je größer die Übereinstimmung der Bewerter hinsichtlich eines Kategoriensystems ist und je mehr die Beobachtungen übereinstimmen (beispielsweise durch den Kappa-Koeffizienten bestimmt), desto eher kann man davon ausgehen, dass eine Kategorie objektiv und reliabel gemessen und geurteilt wird<sup>3</sup>.

Um diese Einflüsse zu reduzieren und eine objektivere Messung spezifischer psychologischer Funktionen und Prozesse zu erhalten, kann in Gedächtnistests etwa die Erinnerungsleistung erfasst werden. Durch Gedächtnistests kann erfasst werden, was Versuchspersonen für wichtiger oder bedeutsamer halten, denn

---

<sup>3</sup>Wobei die Frage offen bleibt, ob denn die Merkmale, anhand derer eine Kategorie bestimmt wird selbst reliable Indikatoren sind.

diese Dinge werden leichter erinnert. Beispielsweise erinnern Versuchspersonen sich eher an Gesichter von Leuten, die als betrügerisch beschrieben wurden im Vergleich zu Personen mit neutralen oder sogar positiven Beschreibungen (Bell u. a.). Bewertende Einstellungen können ebenfalls durch implizite (also indirekte) psychologische Testverfahren erfasst werden, die häufig auf Reaktionszeitmessung basieren. So werden beispielsweise Einstellungen durch *evaluatives Priming* gemessen, bei dem Versuchspersonen außerhalb ihrer bewussten Wahrnehmung („subliminal“) bestimmte Reize vorgegeben werden und danach positive und negative Wörter auf ihre Valenz hin bewertet werden müssen. Hier nimmt man an, dass eine subliminale Präsentation eines negativen Reizes die Bewertung eines negativen Wortes beschleunigt und die Reaktion auf ein positives Wort verlangsamt (siehe Kap. 3).

---

## 2.4 Interne Validität

Wenn wir nun annehmen, dass wir eine oder mehrere reliable Messungen für unsere Konstrukte gefunden haben, dann wollen wir typischerweise auch wissen, wie unsere Konstrukte miteinander kausal in Beziehung stehen. Das heißt, wir wollen herausfinden, ob ein Konstrukt A – als unabhängige oder erklärende Variable – ein anderes Konstrukt B – als abhängige oder erklärte Variable – beeinflusst. Beispielsweise könnte man die Annahme empirisch testen, dass ein bestimmtes Training die Leistung von Piloten verbessert. In einer verlockend einfachen Studie trainiert man eine Gruppe von Piloten mit diesem Training und findet eine Verbesserung ihrer Leistung. Nach Cook und Campbell (1979) könnten hier allerdings einige Dinge schief laufen, die verhindern, dass das Training als erfolgreich interpretiert werden kann. Diese potenziellen Probleme, die verhindern, dass wir Veränderungen in der abhängigen oder erklärten Variablen auch wirklich auf unsere unabhängige Variable zurückführen können, gefährden die interne Validität einer Studie. Was könnte schiefgelaufen sein?

In der Studie könnte sich die Flugleistung alleine dadurch mit der Zeit verbessert haben, dass die Piloten älter werden (etwa bedächtiger), ohne dass das Training eine besondere Wirksamkeit hat. Diese *Alterseffekte* können die abhängige Variable für sich alleine verändern. Außerdem ist es möglich, dass bei der ersten Messung vor dem Training die Piloten besonders nervös waren und deshalb schlechter abgeschnitten haben. Bei der zweiten Messung nach dem Training hätten sie sich schon an die Messung gewöhnt, wären weniger nervös und damit leistungsfähiger. Diese *Testeffekte* meinen alle Einflüsse, die dadurch entstehen, dass Untersuchungsteilnehmer wiederholt mit demselben Test gemessen

werden. Ebenso könnten Piloten einfach mehr Übung mit dem Fliegen in den Testsituationen haben. Diese Praxiseffekte treten insbesondere auf, wenn die Testsituation für die Versuchspersonen neu ist und eine Wiederholung der Messung für sie eine Übung darstellt. Außerdem wissen wir gar nicht, ob bei der Auswahl der Versuchspersonen nicht irgendetwas schief gelaufen ist, sodass nur eine ganz bestimmte Gruppe gewählt wurde, wie beispielsweise alle die, die sich am schnellsten freiwillig gemeldet hatten. Diese *Selektionseffekte* können die Interpretation der Effektivität des Trainings deutlich beeinträchtigen. Auch könnte eine recht schlechte Leistung in der ersten Messung alleine aufgrund der typischen Verteilung, dass mittlere Leistungen rein statistisch wahrscheinlicher sind als extreme, zu dem Eindruck führen, das Training würde wirken. Ebenso könnte dieser Effekt der *Regression zur Mitte* dazu führen, dass eine extrem gute Leistung der Piloten vor Beginn des Trainings zu einem Leistungsabfall bei der zweiten Messung führt, was zur Schlussfolgerung verleiten würde, dass das Training die Flugleistung verringert. Die potenziellen Probleme der internen Validität von Kausalstudien scheinen recht groß zu sein, allerdings kann man sie mit einer recht einfachen Erweiterung der Studie ausräumen. Es ist lediglich notwendig, zu der Gruppe der Versuchspersonen, die eine bestimmte Behandlung bekommen (Experimentalgruppe), noch eine Gruppe hinzu zu nehmen, die keine solche Behandlung bekommt (Kontrollgruppe). Wir haben es dann mit einem experimentellen Versuchsdesign zu tun. Essenziell dafür, dass diese Erweiterung unserer Beispielstudie zum Experiment tatsächlich effektiv die Probleme ausräumt, ist die zufällige Zuordnung der Versuchsteilnehmer auf die beiden Bedingungen (Abb. 2.1).

---

## 2.5 Externe Validität

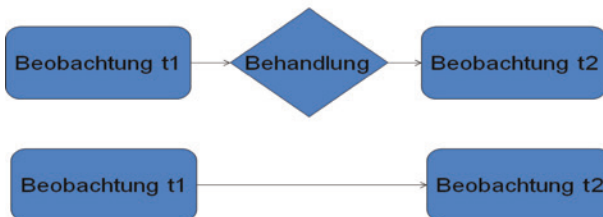
In dem Experimental-Kontrollgruppen Design mit zufälliger oder randomisierter Zuweisung zu den Bedingungen werden also Probleme der internen Validität gelöst. Es bleibt noch das Problem, ob denn die Effekte der Behandlung auch außerhalb des experimentellen Settings einen Einfluss haben, ob sie also auch *extern valide* sind. Auf den ersten Blick scheint dem nichts entgegen zu stehen. Allerdings könnte es sein, dass die Behandlung nur deswegen wirkt, weil die Untersuchungsteilnehmer durch die Messung zum Zeitpunkt 1 auf das Thema der Studie aufmerksam geworden sind oder die erste Messung irgendeinen anderen Einfluss ausübt, der mit der Behandlung interagiert. Um dieses Problem zu lösen, muss die erste Messung weggelassen werden (oder zumindest Bedingungen ohne





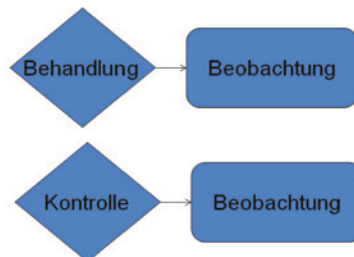
**Probleme der internen Validität:** Zeit- und Alterseffekte, Praxis-Effekte, Test-Effekte, Selektionseffekte, Regression zur Mitte

**Lösung:** Experimental-Kontrollgruppen-Design



**Probleme der externen Validität:** Interaktion der Behandlung mit Beobachtung t1

**Lösung:** Beobachtung t1 weglassen



**Abb. 2.1** Interne und externe Validität, Probleme und Lösung

Vormessung inkludiert werden). Diese Überlegungen begründen das für die Sozialpsychologie typische experimentelle Vorgehen, in dem experimentelle Manipulationen mit ihren Kontrollbedingungen verglichen werden (denen die Probanden zufällig zugeordnet wurden) und danach die abhängige Variable gemessen wird.

Ein Problem ergibt sich selbst in sauber geplanten Experimenten. Es ist das Problem der Interpretation der Ergebnisse. In Experimenten werden Annahmen von Theorien, typischerweise kausale Beziehungen zwischen den Konstrukten, geprüft. Das bedeutet, man kann von Ergebnissen aus Experimenten schließen, ob bestimmte Hypothesen einer Theorie verworfen werden müssen oder die Theorie so beibehalten werden kann. Will man nun beobachtete Ereignisse aus dem persönlichen oder gesellschaftlichen Alltag erklären, müssen die entsprechenden Theorien herangezogen werden, nicht aber bestimmte experimentelle Effekte. Generelle Erklärungen können aus validen Theorien abgeleitet werden und wir prüfen durch Experimente, ob bestimmte Theorien valide sind oder nicht. Generalisierungen und Erklärungen direkt aus Experimenten abzuleiten, wäre fehlgeleiteter Empirismus.

---

## 2.6 Strategien und Probleme, wenn wir Theorien prüfen

Soweit haben wir die Logik des Experiments besprochen, es ist jedoch noch nicht klar, welcher Strategie wir folgen sollten, wenn wir Theorien insgesamt untersuchen wollen. Eine Möglichkeit wäre, so viel Evidenz wie nur möglich für eine Theorie zu sammeln. Je mehr unterstützende Evidenz vorliegt, desto plausibler ist sie. Eine solche Strategie läuft Gefahr, zwei Problemen aufzusitzen. Eines ist im Prinzip lösbar, das andere ist vermutlich nicht wirklich zu lösen (vielleicht praktisch zu ignorieren). Das erste Problem ist die Tendenz, genau die Studien durchzuführen und jene Ergebnisse heranzuziehen, die die eigene Theorie bestätigen. Diese Tendenz ist zu beobachten und vermutlich dadurch zu lösen, dass einerseits die Forscher versuchen, möglichst unvoreingenommen zu sein (was schwierig ist), und andererseits unterschiedliche Forscher und Forschergruppen sich gegenseitig korrigieren, denn was die eine zu bestätigen sucht, kann die andere versuchen zu widerlegen.

Das zweite Problem ist schwieriger wenn nicht gar unlösbar. Es beschreibt das Paradox der Induktion (Hempel 1945), also durch empirische Ergebnisse gesteuerten Wissensgewinn. Es geht davon aus, dass eine Hypothese in all ihren logisch äquivalenten Formulierungen geprüft werden kann und bestätigende Evidenz die Plausibilität der Hypothese erhöht und widerlegende Evidenz die Plausibilität reduziert. Nehmen wir beispielsweise die simple Hypothese „alle Schwäne sind weiß“, hier erhöht jeder weiße Schwan die Plausibilität der Hypothese. „Alle Schwäne sind weiß“ ist auch logisch äquivalent zu dem Satz „Alle nicht-weißen

Dinge, sind keine Schwäne“. Klar!? Logisch würde nun gelten, dass ein gelbes Postauto, als ein nicht-weißes Ding was auch kein Schwan ist, unsere Hypothese „alle Schwäne sind weiß“ bestätigt. Das scheint nicht plausibel. Es gibt noch eine Reihe weiterer Probleme und Paradoxa für die Induktion. Das hat dazu geführt, dass sie mehr und mehr in Verruf kam und nach Alternativen gesucht wurde.

Eine der berühmt gewordenen Lösungen ist die Idee, die Induktion aufzulösen in dem man sie vermeidet. Hypothesen können nicht bestätigt, sondern nur falsifiziert werden (Popper 1934). Um es genauer zu charakterisieren: Existenzhypothesen, die besagen, dass es bestimmte Dinge gibt, können belegt werden in dem man ein Beispiel erbringt. Aber alle Allaussagen – und das sind typischerweise Aussagen über Gesetzmäßigkeiten – können nur falsifiziert werden. Diese Strategie impliziert, dass in gewisser Weise bestätigende Information zwar subjektiv die Plausibilität einer Hypothese erhöhen kann, aber logisch nichts zur Entwicklung von Theorien und zur Wissenserweiterung beiträgt.

In der Forschung werden häufig beide Strategien verwendet. Um nun den Unterschied zwischen beiden Strategien zu verdeutlichen, betrachten wir ein konstruiertes Beispiel: Wenn man die Frustrations-Aggressions-Hypothese, also die Annahme, dass Aggression dann und nur dann auftritt, wenn vorher Frustration erlebt wurde, prüfen möchte, könnte man induktiv vorgehen und Fälle der Frustration beobachten und prüfen, ob denn regelmäßig Aggression folgt. Ebenso müsste man für aggressives Verhalten prüfen, ob denn auch Erlebnisse der Frustration zu finden sind. Man würde also viele Beobachtungen sammeln und aus diesen Beobachtungen die Regelmäßigkeiten generalisieren, dass Aggression aus Frustration folgt. Eine solche Regelmäßigkeit wäre verletzt (oder widerlegt), wenn sich regelmäßig Fälle aggressiven Verhaltens identifizieren lassen, die ohne vorherige Frustration entstehen oder wenn wir Fälle von Frustration beobachten können, aus denen keine Aggression folgt. Das deduktive Vorgehen würde hingegen gezielt auf mögliche Verletzungen der Annahme Aggression folgt Frustration zielen. Hier würde man nach all den Fällen suchen in denen Aggression ohne Frustration oder Frustration ohne Aggression vorkommt.

Beide Strategien des Überprüfens von Theorien, die induktive (aus den empirischen Daten abgeleitet) und die deduktive (aus Theorien abgeleitete) Methode, haben ihre Probleme. Auf der einen Seite ist es schwierig zu bestimmen, was als gute Evidenz für eine Hypothese gelten sollte. Auf der anderen Seite lohnt sich ein strenger Test von Hypothesen insbesondere dann, wenn die Hypothese schon ein gutes Maß an Plausibilität erreicht hat. Ferner benötigen wir induktive Hypothesentestung, da wir häufig keine klare Vorstellung über die Stärke unserer vorhergesagten Effekte haben. Mehrere Studien erlauben uns die Stärke unserer Effekte

einigermaßen zuverlässig abzuschätzen. Jede der Studien sollte dabei die gleichen Konstrukte erfassen, aber etwas unterschiedliche Maße dafür zu verwenden (*konzeptuelle* Replikationen). Dies ist notwendig, wenn einzelne Indikatoren für unsere Konstrukte nur mit einer gewissen (wenn auch hohen) Wahrscheinlichkeit mit diesen Konstrukten verbunden sind. Aus diesen Gründen ist in der Zwischenzeit ein Trend zu induktivem Forschen zu beobachten. Das bedeutet nicht, dass die Falsifikation weniger relevant würde, denn induktives Testen verändert die Wahrscheinlichkeit der Gültigkeit einer Hypothese, aber Hypothesen müssen immer falsifizierbar sein, damit sie überhaupt überprüft werden können.

In psychologischen Studien sollten Hypothesen geprüft werden, die aus Theorien abgeleitet werden, und damit einen Test dieser Theorien darstellen. Das bedeutet, dass wir nicht aus Studien auf den Alltag schließen oder generalisieren. Wenn wir etwas über den Alltag aussagen oder Alltagsphänomene analysieren wollen, müssen wir dies mittels unserer Theorien machen. Es kommt immer wieder vor, dass Studien kritisiert werden, weil es solche „Situationen“ nicht im Alltag gäbe. Eine solche Kritik ist oftmals wenig fundiert, denn die Kritik sollte darauf zielen, ob denn die Theorie wirklich (z. B. in ihrer Allgemeinheit) geprüft wurde. Man nennt den Versuch, Effekte aus Studien auf den Alltag zu extrapolieren „nativen Empirismus“. Ein solcher Versuch schlägt fehl, da bestimmte Studienergebnisse sich immer auf eine bestimmte Umsetzung von Konzepten, bestimmte Stichproben, Kontexte oder Zeitpunkte beziehen. Man müsste unendlich viele Studien machen, um aus diesen irgendwie auf den Alltag zu generalisieren. Wir tun dies nicht. Sozialpsychologische Studien testen Theorien, mittels derer wir Alltagsphänomene genauer verstehen und Handlungsanweisen ableiten.

---

## 2.7 Forschungsstrategien

Es gibt nun eine Reihe unterschiedlicher Herangehensweisen, um Daten für einen Hypothesentest zu sammeln. Welche der Herangehensweisen die vernünftigste ist, hängt davon ab, was genau man untersuchen möchte. Will man den kausalen Einfluss einer Variable auf eine andere untersuchen, dann ist das Experiment die beste Strategie, wenn man am unverstellten Zusammenhang bestimmter Variablen interessiert ist, dann sollte man sich den Strategien der Umfrageforschung bedienen, will man eine Population beschreiben, dann sollte man eine für die Population repräsentative Stichprobe aufsuchen usw.

Berühmt geworden ist eine Umfrage, die das Wahlverhalten der Population der Amerikaner im Jahre 1936 erfassen wollte. Der *Literary Digest* befragte eine Stichprobe von etwa 2 Mio. Teilnehmenden. Dummerweise war die Befragung

eine Telefonumfrage, und im Jahre 1936 hatten nur reiche Bewohner der USA, die typischerweise auch eher Republikaner wählen, ein Telefon. Die Umfrage sagte einen Erdrutschsieg der Republikaner vorher, tatsächlich wurde der demokratische Kandidat Roosevelt gewählt. Was ist genau falsch gelaufen? Die Studie hätte eine repräsentative Stichprobe gebraucht, um herauszufinden, wie das Wahlverhalten der typischen amerikanischen Wähler verteilt ist. Da die Stichprobenziehung verzerrt war, konnte trotz der großen Stichprobe nichts anderes als ein deutlich verzerrtes Ergebnismuster herauskommen. Wann immer wir etwas über eine bestimmte Population aussagen bzw. diese *beschreiben* wollen, müssen wir für diese Population eine repräsentative Stichprobe ziehen. Das ist schwierig, denn meist gelingt es nur hinsichtlich bestimmter Merkmale, die als wesentlich erachtet werden (Alter, Geschlecht, Bildung, Einkommen), eine repräsentative Stichprobe zu ziehen.

### 2.7.1 Korrelationsstudien

Manchmal können wir die vermuteten Kausalvariablen nicht einfach manipulieren (z. B. Alter, Geschlecht), manchmal wäre es ethisch bedenklich (z. B. Bedingungen von Hasskriminalität) und manchmal sind wir zuerst an den Zusammenhängen zwischen bestimmten Variablen interessiert. Unter diesen Bedingungen werden häufig Variablen mittels Fragebögen erhoben und nach Skalenprüfung und -bildung die Zusammenhänge der Variablen geprüft. Dazu sollte auf den interessierenden Merkmalen hinreichend Variation zwischen den Teilnehmenden der Untersuchung bestehen. In Zusammenhangsstudien (aber auch bei Studien, die Kausalität prüfen) geht es typischerweise nicht um die Beschreibung von Populationen, sondern nur darum, vermutete Zusammenhänge empirisch zu prüfen, deswegen ist hier keine repräsentative Stichprobe notwendig. Bei der Interpretation von Korrelationsstudien muss dann natürlich sichergestellt werden, dass nicht etwa über eine bestimmte Population geredet wird, sondern über Zusammenhänge von Variablen oder Konstrukten.

Eine Ausnahme liegt vor, wenn der begründete Verdacht besteht, dass die untersuchten Zusammenhänge nur für bestimmte Populationen gelten könnten. Beispielsweise könnte jemand behaupten, dass ein Gefühl persönlichen Kontrollverlusts Symptome von Depressivität nur in individualistischen Kulturen (nicht aber in kollektivistischen) erhöht. Um diesen Einwand auszuschließen, sollte die verwendete Stichprobe gleichermaßen Mitglieder individualistischer (z. B. Westeuropäer) und kollektivistischer (z. B. Ostasiaten) Kulturen enthalten. Alternativ könnte die Stichprobe in Personen aus individualistischen und kollektivistischen

Kulturen unterteilt und untersucht werden, ob der Effekt in einer dieser Subpopulationen stärker ist als in der anderen. Eine entsprechende Meta-Analyse (also die Betrachtung eines Effekts über mehrere unterschiedliche Studien hinweg; Cheng et al. 2013) zeigt hier übrigens *keinen* Kulturunterschied.

### 2.7.2 Feldstudien

Mit dem Begriff der „Feldstudie“ werden alle Studien bezeichnet, in denen die Teilnehmenden in einem natürlichen Setting, im Unterschied zu einem künstlichen Setting im Labor, untersucht werden. Hierbei kann es sich prinzipiell um alle möglichen Studienformen handeln, wie etwa Beobachtungen, Umfragen, Korrelationsstudien oder auch Experimente (siehe unten). Feldstudien haben den Vorteil, dass die Teilnehmenden in einer eher vertrauten Umgebung befragt werden und nicht durch die Künstlichkeit der Studie beeinflusst werden. Forschende erhoffen sich dadurch eine höhere Übertragbarkeit der Befunde auf Alltagskontexte (externe Validität). Der Nachteil von Feldstudien ist, dass alle möglichen Einflussgrößen eine Rolle spielen können, die in Laborstudien in der Regel kontrolliert werden.

### 2.7.3 Experimente

Um kausale Einflüsse zwischen Variablen zu überprüfen ist das Experiment die Methode der Wahl. Wie wir schon in der Diskussion zu den Bedrohungen der internen Validität deutlich gemacht haben, können alternative Erklärungen eines vermuteten Kausaleffekts dadurch ausgeräumt werden, dass ein Kontrollgruppendesign gewählt wird. Die wesentlichen Merkmale eines Experiments sind (a) die Manipulation der unabhängigen Variablen, (b) die zufällige Zuweisung der Teilnehmenden zu den experimentellen Bedingungen und (c) die „isolierende Bedingungsvariation“. Die *Manipulation der unabhängigen Variablen* (a) besteht beispielsweise darin, dass in der Experimentalgruppe eine bestimmte Intervention (z. B. ein Training zur Toleranz gegenüber Minderheiten) vorgenommen wird, die in der Kontrollgruppe fehlt oder durch eine Intervention ersetzt wird, die die vermuteten Wirkfaktoren (z. B. Kontakt zu Fremdgruppen) nicht enthält. Allerdings ist es gar nicht so einfach, eine bedeutungsvolle Kontrollgruppe herzustellen. Hier muss systematisch vorgegangen werden, um für alle möglichen Merkmale der Intervention eine Kontrolle (einen Vergleich) herzustellen. So könnte ein neu entwickeltes Toleranztraining darauf setzen, dass die Teilnehmenden dieses

Trainings Personen aus Minderheitengruppen (z. B. Zuwanderern) persönlich begegnen (s. Kap. 9). In einer Evaluationsstudie zur Wirkung dieser „Kontaktintervention“ sollte eine Kontrollgruppe sich nur darin von der Interventionsgruppe (d. h., der Experimentalgruppe) unterscheiden, ob Kontakt zu Zuwanderern hergestellt wird oder ob nicht. Alle anderen Merkmale, die den Erfolg einer Intervention alternativ erklären könnten, müssen idealerweise identisch sein. Die Teilnehmenden der Kontrollgruppe sollten also ebenfalls ein Training zum gleichen Thema erhalten, der Kurs sollte den gleichen zeitlichen Umfang haben, von den gleichen Kursleiterinnen durchgeführt werden und an denselben Tagen am gleichen Ort stattfinden. Bloß die Begegnung mit Zuwanderern würde entfallen oder – noch besser – durch eine persönliche Begegnung mit Personen aus der Mehrheitsgesellschaft ersetzt werden. Die *randomisierte Zuweisung* (b) der Untersuchungsteilnehmer bedeutet die zufällige Zuordnung der Versuchspersonen zu den Bedingungen. Sobald irgendwelche Merkmale der Untersuchungsteilnehmer dazu führen, dass Teilnehmende wahrscheinlicher in der einen statt in der anderen Bedingung landen, ist diese wichtige Voraussetzung von Experimenten verletzt. Ein solcher „Randomisierungsfehler“ liegt beispielsweise dann vor, wenn die engagierten Teilnehmenden, die sich zuerst zur Intervention angemeldet haben, von den Organisatoren der Studie eher der Interventionsgruppe zugeordnet werden und die unengagierten Nachzügler eher in der Kontrollgruppe landen. Wenn die Forschenden in einem solchen Fall am Ende der Untersuchung in der Interventionsgruppe positivere Einstellungen gegenüber Zuwanderern finden würden als in der Kontrollgruppe, dann könnten Alternativerklärungen zum Wirken der Kontaktintervention nicht ausgeschlossen werden. Alternative Erklärungen wären dann beispielsweise, dass die Teilnehmenden der Interventionsgruppe von vorn herein fremdenfreundlicher eingestellt oder aber auch empfänglicher für Trainings jeder Art gewesen wären. Schließlich meint die *isolierende Bedingungsvariation* (c), dass man versuchen muss, nicht zu komplexe Manipulationen vorzunehmen, sondern nur den Faktor tatsächlich manipuliert, der laut der Theorie den Einfluss auf die zu erklärende (abhängige) Variable haben soll. In unserem Beispiel würde eine Evaluationsstudie, die sich nur auf den Wirkfaktor „Kontakt zu Zuwanderern“ konzentriert, kausalanalytisch aussagekräftiger sein, als eine, die ein gesamtes Training in all seinen Besonderheiten (z. B. neben dem Kontakt auch Theorieeinheiten zu Stereotypisierung und eine Diskussion über die wirtschaftlichen Vorteile von Zuwanderung) einem Training gegenüberstellt, das all diese Besonderheiten nicht besitzt. Dann ist anschließend nicht klar, inwieweit welche Elemente des Trainings wirklich einen Einstellungswandel bei den Teilnehmenden bewirkt haben.

In der Sozialpsychologie bestehen Experimente häufig aus einem experimentellen Setting, das den Kontext festlegt, innerhalb dessen die Manipulationen auf die abhängigen Variablen wirken sollen. Eine Coverstory und die Instruktionen spezifizieren diesen Kontext genauer, schaffen also für die Teilnehmenden die experimentelle Realität (Aronson et al. 1990), innerhalb derer sie im Verlauf des Experiments denken und handeln sollen (z. B. das Gefühl, unter dem Einfluss einer Autoritätsperson zu stehen, wie in Stanley Milgrams berühmtem Experiment zum Gehorsam; s. Kap. 8). Die Coverstory kann manchmal Informationen liefern, die die Versuchspersonen über den wahren Sinn der Studie täuschen. Dies macht man dann, wenn das Wissen über den wahren Zweck der Studie zu möglichen Urteilsverzerrungen bei den Teilnehmenden führen könnte. Die unabhängige Variable ist die Manipulation, deren Einfluss untersucht werden soll. Die abhängigen Variablen sind jene, für die angenommen wird, dass sie durch die Manipulation beeinflusst werden. Um zu überprüfen, ob die Manipulation tatsächlich den Zustand hervorgerufen hat, den sie hervorrufen sollte, wird häufig eine sog. Manipulation-Check-Variable erfasst (z. B. erlebter Druck durch die Autorität) und auf Unterschiede zwischen den Versuchsbedingungen getestet. Dieser Manipulationscheck ist theoretisch weniger interessant, da hier meist Offensichtliches erfasst wird. Da er aber die Wahrnehmung der Manipulation durch die Untersuchungsteilnehmer bemessen lässt, wird durch ihn zumindest teilweise die subjektive Wirksamkeit des Settings und der unabhängigen Variablen sichergestellt.

Manchmal sind die Experimentatorinnen und Experimentatoren nicht nur an der theoretisch relevanten abhängigen Variablen interessiert, sondern auch an den psychologischen Prozessen, die zur Veränderung in der abhängigen Variable führen. Diese Prozessvariablen (z. B. Indikatoren für Einstellungsänderungen, Verarbeitungstiefe usw.) werden häufig als *Mediatorvariablen* bezeichnet, die genauer erklären, wie die unabhängige Variable auf die abhängige wirkt. Beispielsweise könnten Forschende testen, ob der Effekt von Autoritätseinfluss auf das Handeln der Versuchspersonen durch ein abnehmendes persönliches Verantwortungsgefühl für die eigenen Handlungen vermittelt – und daher erklärt – wird. Schließlich gehört zu den meisten Experimenten eine Reihe von abschließenden Fragen, der postexperimentelle Fragebogen. Hier werden die Teilnehmenden beispielsweise gebeten, über den vermuteten Zweck der Untersuchung Auskunft zu geben. Wenn sie den Zweck der Studie durchschauen, müssen die Ergebnisse manchmal mit Vorsicht interpretiert werden. Schließlich ist dann nicht sichergestellt, dass die Ergebnisse nicht auf offensichtliche Hinweise im Experiment („demand characteristics“) zurückzuführen sind oder auf die Bereitschaft der Versuchspersonen, dem Experimentator einen Gefallen zu tun (oder auch *gegen* den vermuteten Zweck der Studie zu arbeiten).



Am Ende jeder Untersuchung müssen deren Teilnehmende über den tatsächlichen Zweck der Studie aufgeklärt werden. Dies ist insbesondere dann zwingend notwendig, wenn die Studie mit Täuschung (z. B. Cover-Stories die den Zweck der Studie falsch angeben) gearbeitet hat. Hier müssen selbstverständlich auch alle Unwahrheiten aufgeklärt werden.

---

## **2.8      Forschungsethik**

Wir unterscheiden zwei wichtige ethische Aspekte in der Forschung. Auf der einen Seite sind wir als Forschende aufgerufen, uns an bestimmte ethisch-methodische Standards in unserer Forschung zu halten, damit diese tatsächlich zu kumulativer Wissenschaft beitragen kann (Validität und Reliabilität). Auf der anderen Seite arbeiten wir mit Menschen als den Teilnehmenden unserer Untersuchungen. Dies erfordert, dass wir die Rechte unserer Teilnehmerinnen und Teilnehmer unbedingt achten, wie beispielsweise ihre Rechte auf Schädigungsfreiheit und informationelle Selbstbestimmung.

### **2.8.1    Ethische Standards der Forschung**

Wenden wir uns zuerst den ethischen Aspekten valider und reliabler Forschung zu. Es ist selbstverständlich, dass in der Forschung Datenfälschung und Plagiate verboten sind und jene, die so etwas machen schließen sich selbst aus der Wissenschaft aus. Nicht nur für Personen, die aktuell in der Forschung tätig sind sondern auch für Akademikerinnen und Akademiker, die im Berufsleben oder auch in der Öffentlichkeit stehen, beendet solches – auch im Nachhinein festgestelltes – Fehlverhalten Karrieren. Aber nicht nur dieses offensichtliche wissenschaftliche Fehlverhalten, sondern auch unterschiedliche Spielarten schlechter wissenschaftlicher Praxis können die Wissenschaft in die Irre führen. Als wir eingangs unsere Standards besprochen haben, haben wir auf Reliabilität und Validität als zentrale Standards hingewiesen. Beide Aspekte können nur dadurch gewährleistet werden, dass wissenschaftliche Ergebnisse transparent veröffentlicht werden. Zu diesen Veröffentlichungen gehört, dass alle Prozeduren (Messinstrumente, Verfahrensweisen, Setting usw.) genau beschrieben werden, sodass andere Wissenschaftlerinnen und Wissenschaftler die Ergebnisse reproduzieren können.

Für einen echten Test einer Hypothese müssen alle Aspekte der Studie vorher festgelegt sein, wie etwa die Manipulation der unabhängigen Variablen, die Messung der abhängigen Variablen, die Größe der Stichprobe, die statistische

Auswertung, und vieles mehr. Es ist beispielsweise nicht zulässig, eine abhängige Variable mit mehreren Items oder Skalen zu messen und nur die „erfolgreichen“ Messungen zu berichten. Ebenso dürfen Stichproben nicht so lange durch weitere Versuchspersonen ergänzt werden, bis das Ergebnis schließlich statistisch signifikant wird. Dies sind nur zwei Beispiele, wie ein Hypothesentest so weit aufgeweicht werden kann, dass seine tatsächliche Irrtumswahrscheinlichkeit erheblich ansteigt. Für weitere Beispiele sollten Sie den sehr amüsant und eindrucklich geschriebenen Artikel von Simmons et al. (2012) lesen. Wenn solche fehleranfälligen Ergebnisse dennoch den Weg in Publikationen finden, dann kann es zu den gegenwärtig intensiv diskutierten Problemen mangelnder Replizierbarkeit von Forschungsbefunden führen. So ließen sich zahlreiche Ergebnisse, die in der Vergangenheit in renommierten Zeitschriften veröffentlicht wurden in einer umfassenden jüngeren Studie (Nosek und die Open Science Collaboration 2015) nicht replizieren. Nun lässt sich sicherlich schon die Frage danach, was als angemessene Replikation gelten kann, diskutieren (Cumming 2014; Morey et al. 2014). So ist eine direkte Replikation, also die Wiederholung einer Studie mit identischem Versuchsmaterial unter exakt identischen relevanten Rahmenbedingungen, in vielen Feldern der Sozialpsychologie gar nicht möglich. Beispielsweise können sich die Auslöser für Autoritätsgehorsam im Laufe gesellschaftlicher Entwicklung verändern. So ist keineswegs gesichert, dass auch heute – wie zur Zeit des klassischen Milgram-Experiments in den 1960er Jahren – ein Hochschullehrer im weißen Kittel noch einen Auslöser für Autoritätsgehorsam darstellt. Als Alternative werden häufig konzeptuelle Replikationen genutzt, in denen zwar Aufbau (Design) und Variablen einer Originalstudie beibehalten, die Manipulation oder Erfassung der relevanten Variablen allerdings an den jeweiligen (z. B. historischen oder kulturellen) Kontext angepasst werden, in dem eine Untersuchung stattfindet.

Welche Standards können die Gültigkeit und Zuverlässigkeit der Forschung erhöhen? Zum Zeitpunkt des Entstehens diese Lehrbuchs werden in der „scientific community“ eine Reihe von Verbesserungen diskutiert, die sich im Wesentlichen auf Transparenz in Hypothesentestung und Berichterlegung, die Erhöhung von Stichprobengrößen und die Aufwertung von Replikationsstudien beziehen. So wird in sozialpsychologischen Fachzeitschriften mittlerweile größte Transparenz über den Ablauf berichteter Studien verlangt. So sollen Autorinnen und Autoren vollständig berichten, auf welche Weise sie die Stichprobengröße bestimmt haben, ob und welche Daten von der Auswertung ausgeschlossen wurden und welche Manipulationen und Maße in der Studie verwendet wurden (Simmons et al. 2012). Vor typischen Hypothesen testenden Untersuchungen müssen Material und Ablauf der Studie (z. B. die Messung der Variablen) festgelegt sein, wie

auch die analytischen Schritte (z. B. Stichprobengröße, Design der Studie, statistische Tests). Eine mögliche neue Entwicklung könnte sein, dass Studien vor ihrer Durchführung extern durch Fachkolleginnen und -kollegen begutachtet und im positiven Fall bereits dann zur Veröffentlichung angenommen werden, unabhängig von den spezifischen Ergebnissen der Studie. Dies hätte den Vorteil, dass sowohl statistisch signifikante wie nicht-signifikante Ergebnisse veröffentlicht würden. Dadurch könnten Forschende über mehrere Studien hinweg die Einflussstärke der unabhängigen auf die abhängigen Variablen abschätzen. In ähnlicher Weise könnte die sog. Präregistrierung von Studien und deren genaue Dokumentation in öffentlichen Datenrepositorien im Internet (z. B. <https://osf.io>) verhindern, dass im Nachhinein nie gehegte Hypothesen aufgestellt werden oder fehlgeschlagene Hypothesentests im Aktenschränk einer Forscherin oder eines Forschers verschwinden (sog. „file drawer problem“). Welche der gegenwärtig diskutierten konkreten Vorschläge sich in der Forschung endgültig durchsetzen werden, ist nicht vollständig abzusehen. Sie illustrieren jedoch den Kern guter wissenschaftlicher Praxis, die nicht nur auf ein funktionierendes System wissenschaftlicher Selbstkorrektur durch Transparenz und konkurrierende Theorien angewiesen ist, sondern letztlich auf die Pflicht jeder und jedes Forschenden abzielt, die Validität und Reliabilität eigener Forschungsergebnisse zu maximieren und kritisch zu hinterfragen bzw. durch Andere hinterfragen zu lassen.

### 2.8.2 Ethischer Umgang mit Untersuchungsteilnehmenden

Die ethischen Richtlinien sozialpsychologischer Forschung beziehen sich allerdings nicht nur darauf, wie Forschende vorzugehen haben, sondern auch darauf, wie in der Forschung mit den Teilnehmenden der Untersuchungen umgegangen werden muss, sodass ihre Rechte gewahrt bleiben. Zwei ganz grundlegende Regeln dabei sind, dass Teilnehmende selbst über ihre Teilnahme, bzw. die Informationen, die sie preisgeben wollen, bestimmen und dass den Teilnehmenden durch die Untersuchung kein seelischer, körperlicher oder sozialer Schaden entsteht. Obwohl wir Themen behandeln, die für viele Menschen belastend sein können, wie etwa Aggression, Vorurteile und Diskriminierung, dürfen wir das Wohlbefinden von Versuchspersonen nicht über die Untersuchung hinaus beeinträchtigen. Stattdessen sollten Versuchspersonen unsere Studien mit dem Eindruck verlassen, etwas aus der Untersuchung gelernt und zur Beantwortung relevanter Fragen beigetragen zu haben. Tatsächlich sollten Untersuchungsteilnehmende als Partnerinnen und Partner im Forschungsprozess betrachtet werden.

In den meisten Studien ist es angezeigt, eine Einverständniserklärung der Teilnehmenden einzuholen. In diesen Einverständniserklärungen werden Sinn und Zweck der Studie sowie die bevorstehenden Prozeduren erläutert und den Untersuchungsteilnehmern wird klar gemacht, dass die Teilnahme freiwillig ist, die Studie abgebrochen werden kann und dass die Daten anonym und vertraulich und nur zu wissenschaftlichen Zwecken verwendet werden. Sollte es vorhersehbar sein, dass die Teilnehmenden einer Studie Stress, negative Emotionen oder sogar Schmerz verspüren werden, dann muss in der Einverständniserklärung darauf hingewiesen werden.

Eine solche Einverständniserklärung ist dann schwierig zu erhalten, wenn etwa Menschen in ihrem natürlichen Kontext beobachtet werden sollen und der Hinweis, dass sie jetzt gerade an einer Studie teilnehmen ihr Verhalten beeinflussen kann. In solchen Fällen sollten die Beobachteten im Anschluss an die Beobachtung darüber informiert werden, dass sie an einer Studie teilgenommen haben. Natürlich können sie dann die Verwertung ihrer Daten untersagen.

In vielen Fällen werden die Teilnehmenden vor der Studie nicht über deren wahren Zweck informiert. Manchmal wird sogar gezielt eine erfundene Coverstory aufgebaut. Eine solche Täuschung ist nicht immer zu vermeiden. Wenn man beispielsweise Hilfeverhalten untersuchen möchte, darf das nicht vorher schon erwähnt werden, oder sogar darauf hingewiesen werden, dass in der Studie keine echten sondern nur vorgetäuschte Notfälle vorkommen. Dies würde das Verhalten der Versuchspersonen zu sehr beeinflussen, sodass wir aus solchen Studien nicht viel lernen könnten. In den Fällen, in denen Täuschung notwendig ist, müssen die Forschenden sicherstellen, dass die Versuchspersonen im Anschluss an die Studie vollständig aufgeklärt werden und auch hier ihr Einverständnis zurückziehen können. Es wurde untersucht, wie sich Täuschung auf die Teilnehmenden auswirkt. Epley und Huff (1998) fanden beispielsweise, dass Täuschung, wenn sie nicht zu offensichtlich oder verletzend ist, von den Teilnehmenden der Untersuchung nicht als negativ bewertet wird. Es zeigt sich im Gegenteil, dass die Teilnehmenden eher berichten mehr gelernt zu haben, als in den Studien, in denen keine Täuschung vorkam.

---

## 2.9 Zusammenfassung

In diesem Kapitel haben wir die wichtigsten Standards unserer Forschung kennengelernt, wie Validität und Reliabilität. Zu diesen beiden zentralen Standards gehören auch, dass die Forschung öffentlich und kumulativ sein sollte und einfachere Theorien komplexeren vorgezogen werden sollten. Danach wurde auf den

Wert von Theorien hingewiesen, durch die unsere Beobachtungen systematisiert und in einen bedeutungsvollen Zusammenhang gebracht werden. Die Überlegungen zu interner Validität zeigen sehr unterschiedliche Möglichkeiten, wie ein Kausalschluss problematisch sein kann. Das experimentelle Vorgehen behebt mit seinem Experimental-Kontrollgruppen-Design und der zufälligen Aufteilung der Versuchspersonen zu den Bedingungen diese Probleme. Schließlich wurden unterschiedliche Strategien der Überprüfung von Theorien angedeutet und unterschiedliche Forschungsstrategien beschrieben. In einem letzten Abschnitt wurden die ethischen Aspekte der Forschung beschrieben, die einerseits Regeln für die Forschenden angeben (Regeln für gute wissenschaftliche Praxis) und andererseits die Regeln nach denen die Untersuchungsteilnehmer behandelt werden sollten.

---

## Literatur

- Allport, G. W. (1954). The historical background of modern social psychology. In G. Lindzey (Hrsg.), *Handbook of social psychology* (2. Aufl., Bd. 2, S. 3–56). Reading: Addison-Wesley.
- Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzales, M. H. (1990). *Methods of research in social psychology* (2. Aufl.). New York: McGraw-Hill.
- Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193–217.
- Cheng, C., Cheung, S. F., Chio, J. H., & Chan, M. P. (2013). Cultural meaning of perceived control: A meta-analysis of locus of control and psychological symptoms across 18 cultural regions. *Psychological Bulletin*, 139, 152–188.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25, 7–29.
- Davidson, D. H. (2001). *Subjective, intersubjective, objective*. Oxford: Oxford University Press.
- Epley, N., & Huff, C. (1998). Suspicion, affective response, and educational benefit as a result of deception in psychology research. *Personality and Social Psychology Bulletin*, 24, 759–768.
- Hempel, C. G. (1945). Studies in the logic of confirmation (I). *Mind*, 54, 1–26.
- Lewin, K. (1952). *Field theory in social science*. London: Tavistock (hrsg. von Cartwright, D.).
- Morey, R. D., Rouder, J. N., Verhagen, J., & Wagenmakers, E.-J. (2014). Why hypothesis tests are essential for psychological science: A comment on cumming. *Psychological Science*, 25, 1289–1290.
- Nosek & Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac47161–8.
- Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, 12, 1–15.
- Popper, K. (1934). *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien: Springer.

- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *Dialogue*, 26(2), 4–7. <http://ssrn.com/abstract=2160588>.
- Simon, B. (2007). Macht, Identität und Respekt. In B. Simon (Hrsg.), *Macht: Zwischen aktiver Gestaltung und Missbrauch*. Göttingen: Hogrefe.
- The Literary Digest. (November 14, 1936). What went wrong with the polls? *Literary Digest*, 122, 7–8.

Sozialpsychologie

Kessler, Th.; Fritsche, I.

2018, X, 179 S. 26 Abb., Softcover

ISBN: 978-3-531-17126-5