

Grundlagen experimenteller Methoden

Jürgen Bredenkamp

Auf einen der Begründer der experimentellen Psychologie, Wilhelm Wundt, ist eine Definition des Experiments zurückzuführen, die in ihren Grundzügen heute noch akzeptiert wird. In seinem erstmals 1896 erschienenen „Grundriß der Psychologie“ führt Wundt aus: „Das Experiment besteht in einer Beobachtung, die sich mit der willkürlichen Einwirkung des Beobachters auf die Entstehung und den Verlauf der zu beobachtenden Erscheinungen verbindet“ (Wundt, 1920, S. 25). Später finden wir die Aussagen: „Den Inhalt der Psychologie bilden ausschließlich *Vorgänge*, nicht dauernde Objekte. Um den Eintritt und den Verlauf dieser Vorgänge ... exakt zu untersuchen, müssen wir vor allem jenen Eintritt willkürlich herbeiführen und die Bedingungen desselben nach unserer Absicht variieren, was hier wie überall nur auf dem Weg des Experiments geschehen kann“ (Wundt, 1920, S. 26).

Wundt definiert das psychologische Experiment also vom Gegenstand der Psychologie her. Diesen konstituieren für ihn psychische Vorgänge (z.B. „Empfindungen“, „Erinnerungsvorstellungen“, „Gefühle“, „Willensvorgänge“), die durch den willkürlichen Eingriff des Experimentators ausgelöst und verändert werden können und deren Analyse damit nicht vom spontanen Auftreten abhängig ist. Neben diesem Vorteil verband sich für Wundt mit dem Experimentieren die Möglichkeit der „Kausalerklärung“. Viele Psychologen, auch wenn sie Wundts Identifikation der Kausalerklärung mit „der Ableitung komplexerer psychischer Vorgänge aus einfacheren“ (Wundt, 1920, S. 30) nicht nachvollziehen, sind auch heute noch der Auffassung, das psychologische Experiment sei die einzige Methode, die zur Kausalerklärung psychischer Vorgänge führe.

Der Untersuchung psychischer Vorgänge dient nach Wundt also das psychologische Experiment. Der experimentellen Methode steht bei ihm eine zweite Methode der Psychologie gegenüber, nämlich die Beobachtung, die ohne willkürliche Einwirkung „Tatsachen“ untersucht, denen das Merkmal der „relativ beharrlichen Beschaffenheit“ zukommt. „Diese Tatsachen sind die geschichtlich entstandenen *geistigen* Ereignisse, wie die Sprache, die mythologischen Vorstellungen, die Sitten“ (Wundt, 1920, S. 29), deren Analyse er der Völkerpsychologie zuschreibt.

1 Merkmale des Experiments

Wundts Betonung, daß psychische Vorgänge Gegenstand des Experimentierens sind, ist heute noch akzeptabel. Überall in der Psychologie, wo diese untersucht werden – besonders in der sog. Allgemeinen Psychologie, in der u.a. Wahrnehmungs-, Denk- und Gedächtnisprozesse untersucht werden, aber auch in anderen Teilgebieten der Psychologie wie der Sozial- und Entwicklungspsychologie – finden sich experimen-

telle Analysen. Sofern nicht ethische Grenzen dem Experimentieren entgegenstehen (s.u.), ist diese Methode immer dann anwendbar, wenn die Möglichkeit zur willkürlichen Einwirkung und zum systematischen Variieren der Bedingungen besteht. Diese Möglichkeit besteht selbstverständlich nicht immer, etwa dann nicht, wenn der Zusammenhang zwischen dem Lebensalter und der Intelligenztestleistung untersucht werden soll. Ist sie aber gegeben, wird sie in der Psychologie auch genutzt, da die experimentelle Methode wegen der einzigartigen Kontrollierbarkeit von Fehlerquellen (s.u.) am ehesten zu eindeutig interpretierbaren Befunden führt. In vielen psychologischen Fachzeitschriften finden sich deshalb Beispiele für experimentelle Analysen.

Neben der Willkürlichkeit benennt Wundt die Variation der Bedingungen als ein Merkmal des Experiments. Häufig können aber psychische Vorgänge gar nicht unter verschiedenen Bedingungen an ein und demselben Menschen (Probanden) beobachtet werden. Dafür gibt es einen einfachen Grund: Die experimentelle Bedingung *A*, unter der ein Proband (Pb) zunächst steht, beeinflusst den später unter einer Bedingung *B* zu untersuchenden Vorgang häufig mit, so daß der Vergleich der Versuchsergebnisse unter den verschiedenen Bedingungen nicht unbeeinflusst von Übertragungseffekten (*carry-over effects*) von einer Bedingung auf die nächste bleibt. Dieses Problem, das später anhand eines konkreten Experiments verdeutlicht wird (vgl. Abschnitt 4), versucht man in der experimentellen Psychologie häufig so zu lösen, daß verschiedene Probanden unter verschiedenen experimentellen Bedingungen beobachtet werden. Zwei Probanden reichen dafür nicht aus: Sie können sich hinsichtlich bestimmter Persönlichkeitsmerkmale unterscheiden, die außer den unterschiedlichen experimentellen Bedingungen für die beobachteten Unterschiede in den ausgelösten Vorgängen verantwortlich zu machen sind. Die angestrebte Kausalklärung wäre nicht mehr möglich. Deshalb müssen mehrere Probanden unter den verschiedenen Bedingungen beobachtet werden, und ihre Zuweisung zu den Bedingungen wird von dem Experimentator nach dem Zufallsprinzip vorgenommen (Prinzip der Randomisierung). Eine Explikation von Kausalbegriffen, die mit dem randomisierten Experiment in Verbindung gebracht werden können, findet sich bei Steyer (1992).

Für die Randomisierung sind bestimmte Realisierungsvorschriften genau zu beachten, worauf hier nicht einzugehen ist (vgl. dazu etwa Bredenkamp, 1969). Sie soll sicherstellen, daß sich die Untersuchungsgruppen innerhalb kalkulierbarer Fehlergrenzen nur hinsichtlich der experimentellen Bedingungen unterscheiden (stochastische Unabhängigkeit von Gruppenzugehörigkeit und anderen Merkmalen). Von vielen Autoren wird heute die Randomisierung als so wichtig angesehen, daß sie ein weiteres Definitionsmerkmal des Experiments geworden ist. Untersuchungen, in denen zwischen unabhängigen und abhängigen Variablen (s.u.) unterschieden werden kann, eine Randomisierung aber nicht möglich ist, heißen Quasi-Experimente (Campbell & Stanley, 1963).

Die von Wundt genannten Vorgänge müssen sich irgendwie manifestieren, damit sie einer experimentellen Analyse zugänglich werden. Während die experimentellen Bedingungen die sog. unabhängige Variable (UV) eines Experiments konstituieren – es können auch mehrere UVn in eine Untersuchung eingehen – führt die Erfassung der von ihr beeinflussten Vorgänge zur abhängigen Variablen (AV) eines Experiments, wobei auch hier gilt, daß die Beschränkung auf *eine* AV nicht notwendig ist. Infolge der

Randomisierung, die hauptsächlich einer Vermischung unterschiedlicher Ausprägungen von Persönlichkeitsmerkmalen mit den experimentellen Bedingungen vorbeugen soll – häufig wird auch von der Herstellung des Gütemerkmals der internen Validität (Campbell & Stanley, 1963) gesprochen, die die angestrebte Kausalerklärung ermöglichen soll –, entsteht eine Streuung der Werte der AV *innerhalb* der experimentellen Bedingungen. Viele Lehrbücher der experimentellen Versuchsplanung beschäftigen sich aus gutem Grund ausgiebig mit dem Thema, wie diese Streuung zu reduzieren ist. Alle Maßnahmen zur Reduktion dieser Streuung, auf die hier nicht eingegangen werden kann, dienen der Erhöhung der Präzision des Experiments, die neben der internen Validität ein zweites Gütemerkmal psychologischen Experimentierens ist. Der Grund für das Anstreben möglichst hoher Präzision (oder geringer Streuung innerhalb der Bedingungen) ist statistischer Natur. Infolge der zufälligen Zuweisung verschiedener Probanden auf verschiedene Bedingungen kann nicht mehr geprüft werden, wie sich die Variation der Bedingungen auf die psychischen Vorgänge jedes einzelnen Probanden auswirkt; geprüft werden kann nur noch, ob sich die Mittelwerte der AV unter verschiedenen Bedingungen unterscheiden oder ob das nicht der Fall ist. Die Beantwortung dieser Frage nach der Relation zwischen UV und AV wird nicht einer bloßen Inspektion der Daten überlassen, sondern statistisch herbeigeführt. Der Grund dafür ist der, daß unter der Hypothese einer fehlenden Relation (Nullhypothese, H_0) Beobachtungsergebnisse (z.B. Mittelwertsunterschiede in der Untersuchungsstichprobe) auftreten können, die zum Schein für eine Relation zwischen UV und AV sprechen, wie auch unter der Hypothese einer existierenden Relation (Alternativhypothese, H_1) Ergebnisse entstehen können, die fälschlich gegen sie sprechen. Die statistische Beurteilung dient dem Ziel, die Wahrscheinlichkeiten einer fälschlichen Entscheidung für H_1 (sog. α -Fehlerwahrscheinlichkeit) bzw. für H_0 (sog. β -Fehlerwahrscheinlichkeit) gering zu halten (vgl. auch Willmes und Diepgen, in diesem Band). Bei der Durchführung eines sog. Signifikanztests hat es der Experimentator in der Hand, die Wahrscheinlichkeit eines α -Fehlers beliebig klein zu setzen. Allerdings wird die Wahrscheinlichkeit des β -Fehlers umso größer, je geringer α ist. Je größer aber für ein gegebenes α die Präzision eines Experiments ist, umso geringer wird auch die Wahrscheinlichkeit des β -Fehlers. Maßnahmen zur Präzisierung dienen also dem Ziel, eine Relation zwischen UV und AV, falls sie existiert, im Experiment mit größtmöglicher Wahrscheinlichkeit entdecken zu können.

2 Psychologische und statistische Hypothesen

An dieser Stelle ist nun die Frage zu stellen, was die Statistik für die experimentelle Psychologie überhaupt leistet. Randomisiert wird mit dem Ziel der Kontrolle sog. interner Fehler, die die interne Validität eines Experiments und damit auch die angestrebte Kausalerklärung gefährden, und die Maßnahmen zur Präzisierung dienen ebenfalls der Kontrolle, nämlich des β -Fehlers bei der Prüfung der statistischen Hypothese, daß keine Relation zwischen UV und AV besteht (H_0). Für Wundt ist das Experiment die Methode zur Untersuchung psychischer Vorgänge. Diese laufen in einzelnen Menschen ab und sind Gegenstand psychologischer Hypothesen, die sich auf einzelne Menschen beziehen. Sie sind nicht direkt Gegenstand einer statistischen Hypothese (H_0 oder H_1) über die (fehlende) Relation zwischen UV und AV, nach der

die Erwartungswerte der AV unter allen experimentellen Bedingungen gleich oder (in bestimmter Weise) ungleich sind. Die Fehlerkontrolle wird also durch das Prüfen einer statistischen Hypothese erkauft, die mit der psychologisch interessierenden Hypothese nicht identisch ist. Deshalb ist die Überlegung von großer Wichtigkeit, in welchem Verhältnis psychologische (PH) und statistische Hypothesen (SH) zueinander stehen. Die statistische Hypothesenprüfung in der experimentellen Psychologie ist sinnvoll, wenn eine „wenn-dann“-Verknüpfung zwischen PH und SH konstruiert werden kann: „Wenn PH zutrifft, dann gilt auch SH“. In diesem Fall einer implikativen Verknüpfung kann nämlich, falls sich wiederholt zeigen sollte, daß nicht SH, sondern deren Negation gilt, die psychologische Hypothese falsifiziert werden. Bestätigt sich SH, so bewährt sich auch PH. Diese Argumentation macht die Bewährung von SH zur Prüfinstanz für PH. In der obigen „wenn-dann“-Verknüpfung ist SH häufig mit H_1 gleichzusetzen. Man muß dann auf das Zutreffen von „nicht H_1 “, d.h. „ H_0 “, erkennen können, um psychologische Hypothesen falsifizieren zu können. Dazu ist es notwendig, neben der Wahrscheinlichkeit α auch die Wahrscheinlichkeit β zu kontrollieren. Hierzu sollte man sich nicht allein auf Strategien zur Präzisierung eines Experiments verlassen, die die Größe von β noch unbestimmt lassen, sondern von vornherein die statistische Beurteilung so planen, daß α und β Grenzwerte, die der Experimentator festgelegt hat, nicht überschreiten. Die Tabellen Cohens (1988) sowie leicht bedienbare Programme für Personalcomputer (Erdfelder, Faul & Buchner, 1996; siehe auch Buchner, Erdfelder & Faul, in diesem Band) erleichtern die statistische Planung eines Experiments, die die Kontrolle des β -Fehlers a priori vorsieht, erheblich.

Um Empfehlungen bezüglich der festzulegenden Grenzwerte von α und β geben zu können, sind zunächst die Begriffe der Strenge und Fairneß eines Experiments einzuführen (vgl. dazu Erdfelder & Bredenkamp, 1994). Der Grad der Strenge eines Experiments sei $1 - e$, wobei e die Wahrscheinlichkeit dafür ist, daß die PH aufgrund der Daten als bewährt gilt, obwohl diese Hypothese nicht zutrifft. Der Grad der Fairneß sei $1 - f$, wobei f die Wahrscheinlichkeit dafür bezeichnet, daß die PH aufgrund der Daten als nicht bewährt gilt, obwohl die PH zutrifft. Es läßt sich zeigen (vgl. Erdfelder & Bredenkamp, 1994), daß im Falle einer implikativen Verknüpfung von PH und H_1 die Fairneß $1 - \beta$ und die Strenge $(1 - g)(1 - \alpha) + g\beta$ beträgt, wobei α und β die Fehlerwahrscheinlichkeiten des statistischen Tests sind und g die unbekannte Wahrscheinlichkeit dafür ist, daß H_1 zutrifft, obwohl PH falsch ist. (Bei dieser Aussage ist vorauszusetzen, daß die sog. Indifferenzzone hinreichend klein gewählt wurde; vgl. Erdfelder & Bredenkamp, 1994). Am besten geht man bezüglich der statistischen Planung des Experiments so vor, daß zunächst ein minimales Ausmaß an Fairneß festgelegt wird. Unabhängig von der Größe von g läßt sich dann sagen, daß die Strenge der statistischen Prüfung für den festgelegten Wert von β umso größer wird, je kleiner α gewählt wird.

Wird nicht H_1 , sondern H_0 durch die PH impliziert, so gilt (bei vernachlässigbarer Indifferenzzone, s.o.) $1 - e = (1 - g)(1 - \beta) + g\alpha$ und $1 - f = 1 - \alpha$, wobei g die Wahrscheinlichkeit ist, daß H_0 zutrifft, obwohl die PH nicht gilt. In diesem Fall führt die Festlegung von α zur Determination der Fairneß, und für das gegebene α führt jede Reduktion von β zur Erhöhung der Strenge der Prüfung.

Um eine implikative Verknüpfung zwischen PH und H_0 oder H_1 konstruieren zu

können, ist es unerlässlich, das Prinzip der Randomisierung anzuwenden. Ohne die Konstruktion einer derartigen Verknüpfung ist das statistische Hypothesentesten in der experimentellen Psychologie sinnlos, wenn man einmal von dem eher seltenen Fall absieht, daß PH und SH äquivalent sind.

Für die hier dargestellte Argumentation ist vorauszusetzen, daß die psychologische Hypothese als ein sog. Allsatz formuliert wird, der sich auf alle Personen einer offenen Population (nicht unbedingt auf alle Menschen überhaupt) bezieht. Damit ist eine räumlich und zeitlich nicht begrenzte Individuenmenge gemeint (z.B. alle introvertierten Menschen). Eine Begrenzung hätte u.a. zur Folge, daß die Ergebnisse einer Untersuchung zum Zeitpunkt der Veröffentlichung schon veraltet wären, wenn die Population wegen des Todes einiger Mitglieder nicht mehr existiert. Die Formulierung von (deterministischen oder statistischen) Allsätzen ist im übrigen eine Voraussetzung dafür, daß wissenschaftliche Erklärungen geleistet werden können (vgl. Westmeyer, 1973). Wissenschaftliche Erklärungen setzen gut bewährte Hypothesen voraus, und strenge und faire Experimente stehen im Dienste der Erreichung dieses Ziels einer Erfahrungswissenschaft. Ferner muß vorausgesetzt werden, daß die Probanden eines Experiments zu der offenen Population gehören, auf die sich PH bezieht, und daß der methodologische Beschluß akzeptiert wird, bei kleinen Fehlerwahrscheinlichkeiten die Entscheidung gegen die aus der PH abgeleitete SH als ein der psychologischen Hypothese widersprechendes Ergebnis anzuerkennen.

Außerdem ist zu beachten, daß sich psychologische Hypothesen auf Vorgänge beziehen, die sich auf verschiedene Art und Weise manifestieren können. Ihre Erfassung durch *eine* AV – für die UV gilt häufig dasselbe – läßt sich oftmals kaum rechtfertigen, so daß zu fordern ist, einen psychischen Vorgang auf unterschiedliche Art und Weise zu erfassen (konzeptuelle Replikation) und erst dann ein Urteil über PH zu fällen. Eine konzeptuelle Replikation wird meist nur bei Durchführung mehrerer Untersuchungen zu erreichen sein.

Diese Überlegungen (ausführlich dazu Bredenkamp, 1980) zeigen, daß die Falsifikation einer psychologischen Hypothese logisch nicht erzwingbar ist. Der Experimentator kann sich dafür nur mit mehr oder minder guten Gründen *entscheiden*. Alle Planungsschritte laufen darauf hinaus, diese Entscheidung möglichst gut abzusichern und damit ein Korrektiv für die Beliebigkeit des Theoretisierens einzuführen.

3 Einwände gegen das Experiment

Bevor ein Experiment geschildert wird, das einige Aspekte der vorstehenden Ausführungen erläutert, gehen wir, ohne Anspruch auf Vollständigkeit, auf Einwände ein, die häufig gegen das psychologische Experiment vorgetragen werden. Wie wir gesehen haben, werden im Experiment Kontrollen ausgeübt, um die interne Validität und Präzision zu sichern. Wenn es sich um ein hypothesentestendes Experiment handelt – auf diese sog. Entscheidungsexperimente beziehen sich alle vorstehenden Ausführungen – dienen die Kontrollmaßnahmen einer möglichst fehlerarmen Entscheidung über eine Hypothese. Dagegen steht der Einwand, daß die unter der Bedingung strikter Kontrolle erzielten Ergebnisse künstlich und nicht auf die außerexperimentelle Realität übertragbar seien, in der viele Einflußgrößen, die im Experiment „wegkontrolliert“ werden, wirksam sind. Dieser Einwand scheint bei „realitätsnaher“

Durchführung von Untersuchungen die Möglichkeit induktiver Verallgemeinerungen experimenteller Resultate vorauszusetzen, für die es jedoch keine logische Rechtfertigung gibt. Zur Vermeidung des ungelösten Induktionsproblems kann die Frage der praktischen Relevanz eines Experiments, also der Repräsentativität einer Untersuchung für eine bestimmte Praxis, nur mit Hilfe einer Theorie beantwortet werden, die streng und fair geprüft worden und nicht gescheitert ist.

„Die Lösung praktischer Probleme erfolgt hier nicht in der Weise, daß Daten unmittelbar auf die Anwendungssituation übertragen werden, sondern folgendermaßen: Aus bewährten Theorien werden technologische Aussagen abgeleitet. Dies sind Sätze, die besagen, welche Maßnahmen man ergreifen muß, um bestimmte erwünschte Ereignisse eintreten zu lassen, oder um bestimmte unerwünschte Ereignisse zu vermeiden. Die Anwendung einer technologischen Aussage wird durch die Bewährung der Theorien, aus der sie abgeleitet wurde, gerechtfertigt. Die Bewährung dieser Theorien ist das Ergebnis strenger Prüfungen. 'Übertragungen' von Untersuchungsergebnissen auf die Praxis erfolgen also über Theorien. Dies macht eine Übereinstimmung der Untersuchungssituation mit der praktischen Situation überflüssig“ (Gadenne, 1976, S. 80).

Hypothesentestende Experimente machen danach also auch dann, wenn die Lösung praktischer Probleme angestrebt wird, eine strikte Kontrolle notwendig. Geringere Anforderungen bezüglich der Kontrolle sind an ein Erkundungsexperiment zu stellen, das der Generierung von Hypothesen dient.

Ein zweiter Einwand gegen das psychologische Experiment beruft sich darauf, daß dieses eine soziale Situation sei, in der Experimentator und Proband interagieren und die Kommunikation sich nicht allein auf das Verlesen der Versuchsinstruktion und die Reaktion des Probanden auf die vom Experimentator gesetzten Bedingungen bezieht. So soll dieser seine Hypothesen dem Probanden unabsichtlich kundtun, und auf seiten des Probanden werden verschiedene Motivationen wie die, Versuchsleitererwartungen bestätigen zu wollen, in einem günstigen Licht dastehen zu wollen u.a.m., angenommen. Wenn derartige Kognitionen generell experimentelle Ergebnisse verfälschten, bestünde der Einwand gegen das Experiment als valide Methode zur Überprüfung von Theorien zu Recht. Aber von einer generellen Bestätigung dieser „Artefakt“-Hypothese kann keine Rede sein. Sie gilt allenfalls unter sehr speziellen Bedingungen, die keinesfalls typisch für die experimentelle Psychologie sind (ausführlich dazu Bredenkamp, 1980). Dennoch gibt es Autoren, die auf der Grundlage experimenteller Befunde zur „Sozialpsychologie des Experiments“ für eine Abkehr vom Experimentieren in der Psychologie plädieren. Zur empirischen Rechtfertigung des Plädoyers ist es notwendig, die Aussagekraft experimenteller Resultate nur dann nicht in Zweifel zu ziehen, wenn sie im Kontext der „Sozialpsychologie des Experiments“ erzielt worden sind. Die Voraussetzung, die dem Plädoyer zugrunde liegt, ist anzuzweifeln (vgl. Bredenkamp, 1980).

Schließlich sei noch kurz auf den Einwand eingegangen, das Experimentieren mit Menschen sei schlechthin moralisch verwerflich; es sei mit der Menschenwürde nicht vereinbar, da der Proband im Interesse der Erreichung wissenschaftlicher Ziele zu einer „Sache“ werde. Dieser Einwand ist auch auf nicht-experimentelle Untersuchungen mit Menschen übertragbar. Wird aus ihm die Forderung des Verbots psychologischer Versuche abgeleitet, so wird Psychologie als Wissenschaft damit unmöglich gemacht. Tatsächlich wird aber die Menschenwürde in vielen psychologischen Versuchen, auch in experimentellen, nicht, wenn überhaupt, in höherem Maße angetastet als in manchen alltäglichen Situationen auch, in denen etwa Kindern die Auswirkungen von Schulreformen oder durch Umzug der Wechsel sozialer Bezüge zugemutet werden. Selbstverständlich können manche Experimente als moralisch verwerflich beurteilt werden, insbesondere dann, wenn auf eine Täuschung der Probanden, wofür unten ein Beispiel gegeben wird, nicht verzichtet wird. Auf die Täuschung kann insbesondere beim sozialpsychologischen Experimentieren häufig nicht verzichtet werden, aber nicht jedes Experiment, das sich dieser Methode bedient, ist auch schon als moralisch fragwürdig zu bezeichnen. Beachtet werden muß in jedem Fall, daß es nicht zur psychischen oder physischen Schädigung des Individuums kommt. Im übrigen löst man die hier nur kurz angesprochenen Probleme (ausführlich dazu Klauer, 1973) nicht dadurch, daß man auf psychologisches Experimentieren verzichtet. Beispielsweise erscheint die Anwendung bestimmter psychologischer Therapieverfahren, die man nicht der kontrollierten Überprüfung ihrer Wirksamkeit unterzogen hat, als fragwürdig. Nicht nur die Durchführung von Versuchen mit wissenschaftlicher Zielsetzung, sondern auch deren Unterlassung kann ethische Probleme aufwerfen und ist begründungspflichtig.

4 Ein Beispiel: Das Experiment von Festinger und Carlsmith (1959)

Abschließend behandeln wir ein Experiment von Festinger und Carlsmith (1959), das geeignet erscheint, einige Aspekte der vorstehenden Erläuterungen zu erhellen. Untersucht wurden folgende Hypothesen, die zur übergreifenden Theorie der kognitiven Dissonanz gehören: 1) Wird eine Person unter Druck zu Aussagen veranlaßt, die ihrer eigenen Meinung widersprechen, so tendiert sie dazu, ihre Meinung so zu verändern, daß diese im Einklang mit der veranlaßten Aussage steht; 2) diese Tendenz ist umso stärker ausgeprägt, je geringer der angewandte Druck war.

Diese Hypothesen beziehen sich auf den psychologischen Vorgang der Meinungsänderung. Tatsächlich geprüft wurden aber statistische Hypothesen. Um die Übersetzung von PH in SH deutlich werden zu lassen, weichen wir von Festinger und Carlsmith (1959) ab und stellen dar, welche statistischen Hypothesen sie hätten prüfen können. Unter der Voraussetzung, daß die erste Hypothese für jede Person einer offenen Population gilt, läßt sich unter der weiteren Bedingung der Randomisierung die statistische Hypothese ableiten, daß unter Druck die Probanden ein langweiliges Experiment im Durchschnitt interessanter finden als andere Probanden, die ohne Druck ihre Meinung kundtun. Gilt auch die zweite Hypothese für jede Person, folgt bei Randomisierung, daß im Durchschnitt die Pbn unter schwachem Druck ein langweiliges Experiment interessanter finden als andere Probanden, die unter starkem

Druck standen. Der sogleich zu beschreibende Versuchsablauf wird deutlich machen, daß eine wiederholte Erhebung der Meinungen jedes Pb gar nicht möglich war, obwohl bei der psychologischen Hypothesenbildung von einer *Meinungsänderung* die Rede ist.

Drei Versuchsgruppen (ohne Druck, schwacher Druck, starker Druck) nahmen an der Untersuchung teil. Die Pbn (Studenten) dieser Gruppen wurden einzeln untersucht. Zunächst nahmen alle Pbn an einem Leistungsexperiment teil, dessen Ergebnisse die Experimentatoren nicht im geringsten interessierten. Es diente nur dazu, Langeweile und Monotonie bei den Pbn zu induzieren. Nach Beendigung des Experiments wurde ihnen mitgeteilt, sie gehörten der Kontrollgruppe an (Täuschung). Im Unterschied zur Kontrollgruppe werde den Pbn der Experimentalgruppe vor dem Leistungsexperiment gesagt, es sei interessant und fesselnd. Damit es für diese Pbn überzeugend sei, lasse man sie von einem Studenten instruieren. Eben habe der Student, der sonst diese Instruktion vornehme, telefonisch seine Verhinderung bekanntgegeben. Jetzt habe man sich gedacht, daß er, der Pb der Kontrollgruppe, der das Experiment ja kenne, diese Instruktion vornehmen könne. Willigten die Pbn ein, bekamen sie 1 Dollar (schwacher Druck) bzw. 20 Dollar (starker Druck) ausgehändigt. Danach mußte ein Pb der Experimentalgruppe (in Wirklichkeit ein Vertrauter der Experimentatoren) vorschriftsmäßig instruiert werden. Diese sich an das Leistungsexperiment anschließende Instruktion gehörte nicht zu den Aufgaben der Pbn, die keinem Druck ausgesetzt waren.

Die verschiedenen experimentellen Bedingungen (kein Druck, schwacher Druck, starker Druck) werden also durch einen Trick zu einem Zeitpunkt eingeführt, zu dem die Pbn glauben, ihrer Rolle als Pb schon entschlüpft zu sein. Das Täuschungsmanöver dient also dem Ziel der Kontrolle: Die Pbn sollen nicht bemerken, daß (noch) experimentiert wird, damit die Bewußtheit ihrer Rolle als Pb die Versuchsergebnisse nicht verfälscht. Es ist klar, daß derartige Täuschungen nicht mehr als einmal an einem Pb vorgenommen werden können, so daß die Meinungen unterschiedlich behandelter Versuchsgruppen verglichen werden müssen. Die Auszahlung unterschiedlich hoher Geldbeträge soll Druck in unterschiedlichem Ausmaß erzeugen. Hierfür sind noch andere Maßnahmen denkbar, die bei der Überprüfung von dissonanztheoretischen Hypothesen auch realisiert worden sind (konzeptuelle Replikation). Auch das Äußern einer Meinung, die der tatsächlichen Meinung widerspricht, kann auf unterschiedliche Art und Weise erreicht werden. Im Experiment von Festinger und Carlsmith sollten die Pbn ein Experiment, das sie wegen der zu verrichtenden Tätigkeit vermutlich als sehr langweilig erlebt haben, als interessant ausgeben. In den Hypothesen sind theoretische Variablen benannt, die auf unterschiedliche Art und Weise operationalisiert werden können. Bei der Beurteilung eines Experimentalergebnisses ist immer die Möglichkeit zu berücksichtigen, daß die intendierte theoretische UV und AV nicht angemessen operationalisiert wurden, was zur Forderung der konzeptuellen Replikation führt (andere Operationalisierungen). Erst auf der Grundlage mehrerer konzeptuell replizierter Untersuchungen ist eine Hypothese beurteilbar.

Damit die Hypothesen von Festinger und Carlsmith (1959) geprüft werden konnten, war es notwendig, die Meinungen der Pbn zu dem Leistungsexperiment zu erheben. Dies ist wiederum so geschehen, daß die Möglichkeit einer Verfälschung der Ergebnisse infolge bestimmter Motivationen auf seiten der Pbn (s.o.) gering war.

Zum Zweck der Erhebung der AV wurden verschiedene Maßnahmen vorgenommen: Schon bei der Ankündigung der Experimente in einer Vorlesung wurde gesagt, daß eine *andere* Abteilung für Psychologie die Ergebnisse auswerten wolle, um die Versuche verbessern zu können. Deswegen werde eine *Stichprobe* von Teilnehmern interviewt werden, und es sei wichtig, ehrlich zu antworten. Nach Beendigung des oben geschilderten Leistungsexperiments geleitete der Experimentator die Pbn zum Büro des Interviewers und fragte, ob er den Pbn noch sprechen wolle. Der Interviewer, der nicht wußte, welcher Experimentalgruppe der Pb angehörte, bejahte natürlich diese Frage. Alle Maßnahmen dienen dazu, die Pbn den Zweck der Untersuchung nicht erkennen zu lassen.

Die dargestellte Untersuchung kann als prototypisches Beispiel für einen ganzen Bereich der experimentellen Sozialpsychologie gelten. Sie verdeutlicht auch die Wichtigkeit des Randomisierens: Ohne zufällige Zuweisung der Pbn zu den experimentellen Bedingungen bestünde z.B. die Gefahr, daß sich in einer Experimentalgruppe besonders viele änderungsresistente Pbn versammeln, so daß die Unterschiede bezüglich der durchschnittlichen Meinungen nicht auf die UV zurückgeführt werden können (das Experiment wäre in diesem Falle intern nicht valide).

Festinger und Carlsmith (1959) haben mit Studenten experimentiert. Da Studenten zu der Population gehören, für die die Dissonanztheorie Gültigkeit beansprucht, ist diese Beschränkung zulässig. Die Pbn bilden aber keine Zufallsstichprobe aus der in der Theorie gemeinten Population (vgl. Schäffer, in diesem Band). Da diese offen ist, lassen sich Zufallsstichproben auch gar nicht ziehen. Trotzdem sind, sofern randomisiert wurde, statistische Tests zulässig. Das Prinzip der Randomisierung dient nicht nur der internen Validität; es ist auch Voraussetzung für die Durchführung bestimmter statistischer Tests in experimentellen Untersuchungen, in denen keine Zufallsstichproben gebildet werden können (sog. Randomisierungs- oder Permutationstests; vgl. die Kapitel von Bortz und Willmes, in diesem Band).

Die Tatsache, daß keine Zufallsstichproben gezogen wurden, besagt nicht, daß die Pbn, die an der Untersuchung teilgenommen haben, keine repräsentative Stichprobe für die theoretisch gemeinte Population sind. Jeder *einzelne* Pb, sofern er der gemeinten Population zugehört, ist bereits voll repräsentativ für die Untersuchung einer als Allsatz formulierten Hypothese. Nicht aus Gründen der Repräsentativitätserhöhung, sondern der Fehlerkontrolle wird mit mehr als einem Pb experimentiert (s.o.). Dies gilt auch dann, wenn die wiederholte Beobachtung eines Pb unter verschiedenen Bedingungen möglich ist. In diesem Falle sollte die Reihenfolge der experimentellen Bedingungen je Pb randomisiert werden, sofern nicht die o.g. Übertragungseffekte Gegenstand der Untersuchung sind.

Das Experiment von Festinger und Carlsmith ist „künstlich“; seine Ergebnisse lassen sich nicht induktiv verallgemeinern. Es ist ein Prüfungsexperiment für eine Theorie, die die Ableitung technologischer Aussagen zuläßt.

5 Weiterführende Literatur

Das Buch von Bredenkamp (1980) stellt in Verbindung mit dem Artikel von Erdfelder und Bredenkamp (1994) ausführlich die hier nur kurz dargestellten bzw. nicht berücksichtigten Aspekte aus der Sicht einer deduktivistischen Theorie des Expe-

riments dar. Eine induktivistische Auffassung vom Experiment vertreten Campbell und Stanley (1963). Ihr Artikel ist ein „Klassiker“, der verschiedene experimentelle und quasi-experimentelle Versuchspläne behandelt. Im Buch von Gadenne (1976) findet sich eine Kritik der induktivistischen Theorie experimenteller Gütekriterien nach Campbell und Stanley (1963).

Zur Versuchsplanung und -auswertung gibt es viele Lehrbücher. Als deutschsprachige Literatur werden besonders Hager (1992) sowie Henning und Muthig (1979) empfohlen. Eine Übersicht über experimentelle Kontrolltechniken und die damit verbundenen Probleme bietet z.B. Bredenkamp (1969).

Literaturverzeichnis

- Bredenkamp, J. (1969). Experiment und Feldexperiment. In C. F. Graumann (Hrsg.), *Sozialpsychologie* (= Handbuch der Psychologie, Band 7, 1. Halbband: Theorien und Methoden, S. 332–374). Göttingen: Hogrefe.
- Bredenkamp, J. (1980). *Theorie und Planung psychologischer Experimente*. Darmstadt: Steinkopff.
- Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago: Rand Mc Nally. (Deutsche Bearbeitung: Schwartz, E. (1970). Experimentelle und quasi-experimentelle Anordnungen in der Unterrichtsforschung. In K. Ingenkamp & E. Parey (Hrsg.), *Handbuch der Unterrichtsforschung*. Teil 1. Weinheim: Beltz.)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale: Erlbaum.
- Erdfelder, E. & Bredenkamp, J. (1994). Hypothesenprüfung. In T. Herrmann & W. Tack (Hrsg.). *Methodologische Grundlagen der Psychologie* (= Enzyklopädie der Psychologie, Themenbereich B, Serie I, Band 1, S. 604–648). Göttingen: Hogrefe.
- Erdfelder, E., Faul, F. & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1–11.
- Festinger, L. & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203–210.
- Gadenne, V. (1976). *Die Gültigkeit psychologischer Untersuchungen*. Stuttgart: Kohlhammer.
- Hager, W. (1992). *Jenseits von Experiment und Quasi-Experiment. Zur Struktur psychologischer Versuche und zur Ableitung von Vorhersagen*. Göttingen: Hogrefe.
- Henning, H. J. & Muthig, K. (1979). *Grundlagen konstruktiver Versuchsplanung*. München: Kösel.
- Klauer, K. J. (1973). *Das Experiment in der pädagogischen Forschung*. Düsseldorf: Schwann.
- Steyer, R. (1992). *Theorie kausaler Regressionsmodelle*. Stuttgart: Fischer.
- Westmeyer, H. (1973). *Kritik der psychologischen Unvernunft*. Stuttgart: Kohlhammer.
- Wundt, W. (1920). *Grundriß der Psychologie* (14. Aufl.). Stuttgart: Kröner.