

Sebastian Jäckle

Lange Zeit pflegte die Politikwissenschaft ein ambivalentes Verhältnis zur experimentellen Methode. Experimente wurden vielfach als diejenige Vorgehensweise betrachtet, die dem Ideal wissenschaftlich erklärender Erkenntnis am nächsten komme (Lijphart 1971, S. 683) bzw. die „Urform aller Untersuchungsdesigns“ (Behnke et al. 2010, S. 49) sei, da sie insbesondere für die Überprüfung von Kausalitäten geeignet seien. Andererseits wurde im selben Zug zumeist darauf hingewiesen, dass eine Anwendung von Experimenten in der Politikwissenschaft aus ethischen, praktischen oder forschungslogischen Gründen in aller Regel nicht möglich sei (Aarebrot und Bakka 2003, S. 62; Lijphart 1971, S. 49–50). Bereits im Jahr 1910 verweist Lowell auf die begrenzten Untersuchungsoptionen, die der Politikwissenschaft zur Verfügung stehen, da es für sie unmöglich sei auf Experimente zurückzugreifen. Hieraus leitet er sein in den folgenden Jahrzehnten so gut wie nie hinterfragtes Diktum ab: „politics is an observational, not an experimental science“ (Lowell 1910, S. 7).¹ Gleichwohl wird in Lehrbüchern oftmals gefordert, dass man mit den der Politikwissenschaft zur Verfügung stehenden Methoden versuchen sollte, dem Ideal experimenteller Forschung möglichst nahe zu kommen. Daher

¹ Ein frühes Beispiel, dass Experimente durchaus auch in der Politikwissenschaft zielführend eingesetzt werden können, liefert Harold Gosnell (1926). Er zeigte mit einem großangelegten Experiment, dass das gezielte Anschreiben von Wahlberechtigten mit Informationen wie man sich zur Bürgermeisterwahl in Chicago registrieren kann, Auswirkungen auf die Wahlbeteiligung hat. Diese Studie kann damit als einer der wenigen Vorläufer für die seit Anfang der 2000er Jahre verstärkte Hinwendung zu Feldexperimenten gelten. Zum Teil widmen diese sich sogar ebenfalls der Wählermobilisierung, wie Gosnell fast eine Dekade zuvor (z. B. Gerber und Green 2000).

verwundert es auch nicht, dass politikwissenschaftliche Vorgehensweisen vielfach (wie z. B. bei Lijphart 1971, S. 683) im Vergleich zum Experiment diskutiert werden: „Die Beurteilung eines Untersuchungsdesigns lässt sich entsprechend unter dem Aspekt vornehmen, inwieweit es die Stärken des experimentellen Designs besitzt, oder seine Schwächen auszugleichen versteht“ (Behnke et al. 2010, S. 50). Je nach Schule wurden inferenzstatistische wie auch qualitativ vergleichende Analyseverfahren als bestmögliche Annäherungen an die experimentelle Methode gesehen. Gleichzeitig wurden jedoch lange Zeit wenige Versuche unternommen wirklich experimentell zu arbeiten. Seit Ende der 1990er Jahre hat sich dies gewandelt. Seither fanden zunächst v. a. im angelsächsischen Sprachraum vermehrt Experimente ihren Weg in politikwissenschaftliche Zeitschriften (Druckman et al. 2006; Morton und Williams 2008, S. 339–340).² 2010 wurde innerhalb der APSA die Sektion *Experimental Research* gegründet³ und auch die in den letzten Jahren vermehrt zu findenden Lehrbücher, Sonderhefte und Beiträge in Sammelbänden zu Experimenten zeugen von wachsendem Interesse an dieser Methode insbesondere innerhalb der englischsprachigen Politikwissenschaft (Druckman et al. 2011; Gerber und Green 2009; Morton und Williams 2008). Mittlerweile wird aber auch in Deutschland dem „unbekannten Wesen“ (Faas 2009) Experiment bereits der Aufstieg vom „Mauerblümchen zum Mainstream“ (Faas und Huber 2010) attestiert. Zurückführen lässt sich diese positive Entwicklung auf zwei Aspekte: 1) den informationstechnologischen Fortschritt, der das kostengünstige Durchführen von Experimenten erst ermöglicht hat;⁴ 2) ein seit den 1990er Jahren innerhalb der

² Zudem weisen Druckman et al. (2006, S. 628–629) darauf hin, dass ein Wandel im Sprachgebrauch stattgefunden hat. Wurde der Begriff Experiment in früheren Jahren (bis in die 1990er hinein) entweder im Sinne eines institutionellen Novums (z. B. das Experiment eines neuen Wahlsystems) oder im Sinne eines vergleichsweise unsystematischen „Ausprobierens“ verwendet, so begreift die jüngere Literatur unter einem Experiment eine klar umrissene, eindeutig definierte Methodologie. Ähnliche Ergebnisse finden Faas und Huber (2010, S. 723–724) für den deutschsprachigen Raum. Ihrer Analyse zufolge verwenden viele Autoren hier das Wort Experiment auch in jüngerer Zeit (ihr Untersuchungszeitraum beginnt im Jahr 2000) noch im Sinne von „ausprobieren“ wohingegen Experimente im eigentlichen Sinn nur sehr vereinzelt in den Zeitschriften anzutreffen sind.

³ Diese gibt mit dem *Journal of Experimental Political Science* auch eine Zeitschrift heraus, die sich ausschließlich der Anwendung und methodischen Weiterentwicklung experimenteller Verfahren in der Politikwissenschaft verschrieben hat. Daneben existiert mit *Political Methods: Experiments & Experimental Design* ein von James Druckman herausgegebenes SSRN E-Journal, in dem ebenfalls praktische Anwendungen und Methodendiskussionen von Experimenten publiziert werden.

⁴ Heute gibt es an vielen Universitäten speziell eingerichtete Experimentallabore. Organisatorisch sind diese freilich eher bei der Ökonomie oder der Psychologie denn bei der Politikwissenschaft angesiedelt, gleichwohl können oftmals auch Forschende und Studierende aus

Politikwissenschaft verstärkt aufkommendes Bewusstsein für die Wichtigkeit von Kausalitäten für die Erklärung politischer Phänomene (King et al. 1994).⁵

Im folgenden Abschnitt wird zunächst kurz die Grundlogik experimenteller Designs vorgestellt, aus der hervorgeht, was die Vorteile dieser Form des methodischen Vorgehens gegenüber anderen Untersuchungsdesigns gerade in Bezug auf die kausale Erklärungskraft sind. Wir werden sehen, wo die Schwierigkeiten in der Anwendung gerade in der Politikwissenschaft liegen, aber auch welche Möglichkeiten es trotzdem auch für Master-Studierende gibt experimentell zu arbeiten. Daneben wird diskutiert, wann Experimente unter Laborbedingungen und wann besser im Feld durchgeführt werden und welche Konsequenzen Abschwächungen der experimentellen Bedingungen im Sinne von Natürlichen Experimenten und Quasi-Experimenten auf die Durchführbarkeit und die Aussagekraft der Ergebnisse haben.

der Politikwissenschaft diese als Fachfremde nutzen. Teilweise bieten Experimentallabore umfassende Dienstleistungen an. So können Forscher, die das Hamburger Wiso-Forschungslabor zur Benutzung buchen, auch direkt auf einen „Pool von potentiellen, bei Bedarf jederzeit abrufbaren Teilnehmerinnen und Teilnehmer[n]“ zurückgreifen (<http://www.wiso.uni-hamburg.de/forschung/forschungslabor/experimentallabor/buchung/>) und werden von Mitarbeitern des Experimentallabors bei forschungspraktischen Fragen beraten. Die Entwicklungen im Bereich des Internet haben außerdem dazu geführt, dass viele Experimente heute bereits komplett online durchgeführt werden können: Eine Option, die insbesondere für Master-Studierende aufgrund der vergleichsweise einfachen Durchführbarkeit interessant sein dürfte. Spezifische Probleme der externen Validität insbesondere bei der Verwendung von Studierenden als Teilnehmer von Experimenten sowie bei Online-Experimenten werden weiter unten noch genauer beleuchtet.

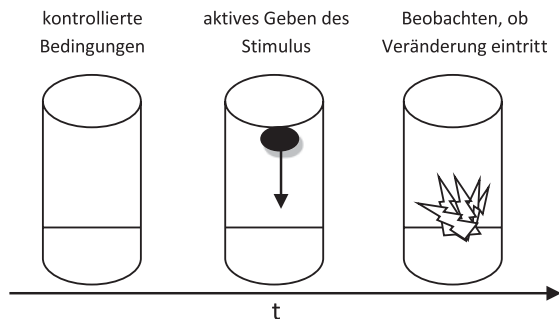
⁵ King et al. halten Experimente allerdings in den Sozialwissenschaften für zumeist nicht anwendbar (1994, S. 188, 197). Sie propagieren stattdessen systematische Vergleichsanlagen, mit deren Hilfe man kausale Inferenz zumindest approximieren könne. Das experimentelle Design ist in ihren Augen vor allem wichtig, um aus ihm heraus Erkenntnisse über nicht-experimentelle Forschungsanlagen zu gewinnen (King et al. 1994, S. 125). In Abgrenzung zu King und dessen Kollegen zeigen George und Bennett (2005) mit der Methode des Process Tracing bei Einzelfallanalysen eine andere Möglichkeit auf, Fragen der Kausalität anzugehen. Die von King, Keohane und Verba angebotenen Antworten sind darüber hinaus heftig diskutiert und kritisiert worden (Brady und Collier 2010).

2.1 Grundlogik experimenteller Designs – das Beispiel mit dem Reagenzglas

Das Ziel eines Experiments ist die kausale Erklärung eines Phänomens. Es geht damit in seinem Erkenntnisinteresse über die Feststellung eines rein korrelativen Zusammenhangs hinaus. Die Frage ist also nicht nur, ob X und Y zusammenhängen, sondern ob sich entsprechend einer gerichteten Hypothese eine unabhängige Variable (UV) X kausal auf die abhängige Variable (AV) Y auswirkt. Im Experiment lassen sich sowohl einfache bivariate wie auch konvergente Kausalstrukturen mit mehreren UV oder komplexere Pfadmodelle testen. Ursprünglich stammt die experimentelle Vorgehensweise aus den Naturwissenschaften. Gießt ein Chemiker beispielsweise einige Milliliter Brom in ein Reagenzglas und gibt dann eine kleine Kugel aus Alufolie hinzu, so wäre dies eine einfache experimentelle Versuchsanordnung. Nach einigen Sekunden beobachtet er, dass die Aluminiumkugel anfängt, mit einer hellen Flamme und unter starker Hitzeentwicklung zu verbrennen. Der Chemiker folgert daraus, dass Aluminium mit Brom exotherm reagiert. Zwei Aspekte sind es, die diesen Versuch als Experiment kennzeichnen: 1) die Kontrolle sämtlicher *Drittfaktoren* (auch *Störfaktoren* genannt), die potentiell einen Einfluss auf das Phänomen haben könnten, und 2) die gezielte Isolation eines ursächlichen Faktors dadurch, dass der Forscher aktiv diejenige UV manipuliert, von der er einen Einfluss erwartet. Man spricht in diesem Zusammenhang auch davon, einen *Stimulus* oder ein *Treatment* zu geben.

Im Chemie-Beispiel sind beide Kriterien erfüllt (vgl. Abb. 2.1). Es handelt sich um ein Laborexperiment, bei dem sichergestellt werden kann, dass sich ausschließlich Brom und keine weiteren Substanzen im Reagenzglas befinden. Der Forscher gibt selbst das Aluminium hinzu und setzt damit aktiv den Stimulus. Weitere ursächliche Faktoren können aufgrund der eindeutig spezifizierten Versuchsanordnung ausgeschlossen werden. D. h. sollte sich nach der Zugabe der Aluminium-

Abb. 2.1 Beispiel für experimentelles Vorgehen



kugel eine Veränderung innerhalb des Reagenzglases beobachten lassen, so muss diese auf das Treatment zurückzuführen sein. Durch die klare Charakterisierung der Untersuchungsanordnung und des Untersuchungsverlaufs lässt sich das Experiment auch beliebig häufig von unterschiedlichen Forschern replizieren. Objektivität und Reliabilität der Ergebnisse können so gewährleistet werden.

2.2 Experimente in den Sozialwissenschaften

Auch in den Sozialwissenschaften müssen bei einem Experiment Drittfaktoren kontrolliert, eine als potentiell relevant erachtete UV aktiv durch den Forscher manipuliert und letztlich das interessierende Phänomen in dieser so eindeutig gekennzeichneten Situation beobachtet werden. Da soziale und politische Phänomene jedoch nicht im Reagenzglas stattfinden, sondern in der Politikwissenschaft unter anderem politische Institutionen, Parteien, makroökonomische Faktoren und vielfach Individuen betrachtet werden, stellen sowohl die Kontrolle von Störvariablen⁶ als auch die Stimulusabgabe den Experimentator oftmals vor große Probleme. Wie solche Störvariablen kontrolliert werden können, zeigt der folgende Abschnitt anhand eines Beispiels: Es soll der Frage nachgegangen werden, inwiefern Wahlkampfspots Auswirkungen auf die Wahlentscheidung haben. Um dies zu untersuchen, zeigt man Probanden eine Fernseh-Wahlwerbung und fragt sie im Anschluss nach ihrer Wahlabsicht.

Um aus diesem Versuch ein experimentelles Forschungsdesign zu machen, sind jedoch einige Aspekte zu beachten. Zunächst müssen mögliche Drittvariablen kontrolliert werden, die ebenfalls einen Einfluss auf die Wahlabsicht haben könnten. Besonders offensichtlich ist dies bei den personenbezogenen Merkmalen. Die Probanden haben (höchstwahrscheinlich) bereits vor dem Betrachten der Wahlkampfspots eine gewisse politische Überzeugung und eine entsprechende Wahlabsicht. Zudem wissen wir, dass es bestimmte soziodemographische Merkmale gibt, wie beispielsweise das Alter oder die Schichtzugehörigkeit, die sich ebenfalls auf die Wahlabsicht auswirken. Daneben kann es noch weitere Faktoren geben, die die Wahlabsicht beeinflussen, von deren Existenz man als Forscherin eventuell noch gar nichts ahnt.

⁶ Vier Arten von Störfaktoren können unterschieden werden: 1) Situationsmerkmale (In welcher Situation findet das Experiment statt? Ist diese Situation für alle Probanden dieselbe?); 2) Versuchspersonenmerkmale (z. B. Alter, Geschlecht, Bildungsgrad der Probanden); 3) Versuchsleitermerkmale (z. B. Art des Auftretens, Alter, Geschlecht des Experimentators); 4) Sequenzeffekte bei wiederholter Messung (z. B. Ermüdungs- oder Lerneffekte, oder zwischenzeitliches Geschehen, welches einen zusätzlichen Einfluss neben dem Stimulus auf die AV hat).

2.3 Experimental- und Kontrollgruppe

Da es nicht möglich ist, die Untersuchungspersonen, wie in dem Chemie-Experiment, bei dem außer Brom und Aluminium keine weiteren Substanzen in das Reagenzglas gelangten, gewissermaßen von ihren die Wahlabsicht beeinflussen den Persönlichkeitsmerkmalen zu trennen, behilft man sich durch einen Kniff. Man teilt die Probanden in eine Test- und eine Kontrollgruppe auf. Die beiden Gruppen sollten sich im Idealfall ausschließlich dadurch unterscheiden, dass der Experimentalgruppe der Stimulus gegeben wird, wohingegen die Kontrollgruppe keinen Stimulus erhält. In unserem Beispiel würde der Experimentalgruppe beispielsweise ein Wahlkampfspot der SPD gezeigt, während der Kontrollgruppe ein etwa gleichlanger, nicht mit (Partei)Politik konnotierter Film, z. B. ein Werbespot für ein Putzmittel, vorgeführt wird. Zeigt sich nun ausschließlich in der Experimentalgruppe ein Effekt, d. h. geben nach Betrachten des SPD-Spots mehr (oder auch weniger) Personen als vor dem Betrachten des Spots an, die SPD wählen zu wollen, muss dieser Effekt – da ansonsten ja keine Unterschiede zwischen den beiden Gruppen vorliegen – auf das Treatment zurückzuführen sein. Ein solches einfaches experimentelles Design wird Vorher-Nachher-Design oder auch Pre-Test-Post-Test-Design genannt. Tabelle 2.1 präsentiert es in der klassischerweise für die Darstellung von experimentellen Designs verwendeten Notation. Weitere Varianten experimenteller Designs, mit deren Hilfe auch nicht-dichotome abhängige Variablen oder auf Wechselwirkungen zwischen vorheriger Messung und Treatment getestet werden können zeigen Behnke et al. (2010, S. 63–65).

Wie teilt man die Probanden nun aber auf die Experimental- und Kontrollgruppe auf, sodass sich diese möglichst nicht unterscheiden? Hierzu gibt es grundsätzlich zwei mögliche Verfahren: Das Matching sowie die Randomisierung.

2.3.1 Matching

Für das Matching benötigt man zumindest eine begründete Vermutung über mögliche Störvariablen. Die Verteilung der Probanden in den beiden Gruppen wird dann entsprechend der Ausprägungen dieser potentiellen Störfaktoren angegli-

Tab. 2.1 Vorher-Nachher-Design

Zeitpunkt	t ₁	t ₂	t ₃
Experimentalgruppe	O	X	O
Kontrollgruppe	O		O

X Treatment/Stimulus, O Observation/Beobachtung

Tab. 2.2 Zuweisung zu Experimental- und Kontrollgruppe durch parallelisierte Paare

Probandenpaar	Experimentalgruppe	Kontrollgruppe
1	18–35, ♂, Abitur	18–35, ♂, Abitur
2	36–60, ♀, Studium	36–60, ♀, Studium
3	über 60, ♀, Hauptschule	über 60, ♀, Hauptschule
4	36–60, ♂, Hauptschule	36–60, ♂, Hauptschule
⋮	⋮	⋮

chen. Zwei Arten des Matchings können unterschieden werden: Die Methode der parallelisierten Paare und die der parallelisierten Gruppen (vgl. Behnke et al. 2010, S. 49–50). Bei ersterer werden Paare gebildet, die sich bezüglich der angenommenen Störvariablen nicht unterscheiden. Jeweils eine Person eines solchen Paares wird dann zufällig der Experimental- die andere der Kontrollgruppe zugewiesen. Damit sind die beiden Gruppen hinsichtlich der erwarteten Störfaktoren vollkommen identisch (vgl. Tab. 2.2). Problematisch wird diese Vorgehensweise bei einer großen Anzahl möglicher Störvariablen und/oder wenn diese viele Ausprägungen aufweisen. Beides erhöht die Zahl an Kombinationsmöglichkeiten, wodurch es immer komplizierter wird in Bezug auf die Ausprägung der Störfaktoren zwei komplett identische Probanden zu finden.

Die Methode der parallelisierten Gruppen weicht deshalb die Vorgaben auf: einzig die Verteilung der Probanden bezüglich der einzelnen Störvariablen muss dabei nun in Experimental- und Kontrollgruppe identisch sein. Die einzelnen Personen müssen dies jedoch – mit ihren jeweiligen Kombinationen von Störfaktorausprägungen – nicht sein. Angenommen, das Experiment soll mit insgesamt 200 volljährigen Probanden durchgeführt werden, die nach ihrem Geschlecht, nach drei Altersstufen (18–35, 36–60, über 60) und nach drei Bildungsabschlüssen (maximal Hauptschule, Abitur, Studium) kategorisiert werden. Die Zuweisung zu den beiden Gruppen erfolgt dann so, dass in den beiden Gruppen dasselbe Verhältnis für alle drei Variablen vorherrscht wie in der Stichprobe aller Probanden. So sollten bei einer Gesamtzahl von 110 Frauen jeweils 55 in der Experimental- und 55 in der Kontrollgruppe sein (vgl. Tab. 2.3).

Letztlich gilt es bei der Wahl der Matching-Methode abzuwägen zwischen einem möglichst guten Ergebnis, d. h. einer möglichst identischen Verteilung der Störfaktoren in Kontroll- und Experimentalgruppe (parallelisierte Paare) und der praktischen Machbarkeit (parallelisierte Gruppen).

Tab. 2.3 Zuweisung zu Experimental- und Kontrollgruppe durch parallelisierte Gruppen

Störvariable	Ausprägung	Gesamt	Experimental- gruppe	Kontrollgruppe
Geschlecht	♀:	110	55	55
	♂:	90	45	45
Alter	18–35:	40	20	20
	36–60:	110	55	55
	über 60:	50	25	25
Bildung	Hauptschule:	30	15	15
	Abitur:	90	45	45
	Studium:	80	40	40

2.3.2 Randomisierung

Oftmals ist im Vorhinein allerdings nicht klar auszumachen, welche Faktoren sich neben dem eigentlichen Stimulus noch auf das Ergebnis auswirken könnten. Für diese Störvariablen kann dementsprechend auch nicht gezielt eine Gleichverteilung in den beiden Gruppen generiert werden. Aus diesem Grund behilft man sich über eine Zufallsauswahl (Faas und Huber 2010, S. 725; McDermott 2002, S. 33; Morton und Williams 2008, S. 342–343). In unserem Fall wird entsprechend zufällig entschieden – beispielsweise durch Münzwurf – ob ein Proband den Wahlkampfspot oder die Putzmittelwerbung vorgeführt bekommt. Die verbliebenen Differenzen zwischen den Gruppen beruhen dann ausschließlich auf einem Zufallsfehler. Dieser wird kleiner, je mehr Probanden man zufällig den beiden Gruppen zuweist. Die Randomisierung führt mit wachsender Gruppengröße deshalb dazu, dass sich Experimental- und Kontrollgruppe entsprechend der Verteilung sämtlicher Eigenschaften – also auch der möglichen Störfaktoren – angleichen. Bei sehr kleinen Gruppen würden einzelne, deutlich abweichende Fälle eine systematische Verzerrung zwischen den beiden Gruppen bewirken. Je mehr Probanden in den Gruppen sind, desto eher gleichen sich solche Fälle aus. Ab etwa 30–40 Personen in jeder Gruppe kann der Forscher relativ sicher sein, dass eine rein randomisierte Zuweisung zwei annähernd identische Gruppen generiert (Behnke et al. 2010, S. 61). Zudem ist es natürlich möglich mehr als eine Treatment-Gruppe zu bilden (Gerring 2012, S. 258–259), wodurch in unserem Fall beispielsweise der Effekt unterschiedlicher Wahlkampfspots miteinander verglichen werden könnte. Häufig müssen Experimente jedoch aufgrund forschungspraktischer oder finanzieller Gründe mit deutlich weniger Probanden auskommen. In einem solchen Fall ist es von Vorteil, wenn mittels Matching einzelne Störvariablen gezielt ausgeschaltet und die weiteren nicht bekannten Störvariablen über eine anschließend durchgeführte randomisierte Zuweisung der Probanden zu den beiden Gruppen kontrolliert werden können.

2.4 Elimination und Konstanthalten von Störfaktoren

Potentielle, das Ergebnis beeinflussende Merkmale der Probanden lassen sich also über eine Aufteilung in Experimental- und Kontrollgruppe kontrollieren. Für Störvariablen, die mit der Untersuchungssituation oder dem Versuchsleiter zusammenhängen, muss anders vorgegangen werden. Zum Teil lassen sich solche Störvariablen komplett ausschalten. Insbesondere in Laborexperimenten⁷ ist eine *Elimination* einzelner, potentiell das Ergebnis verzerrender Faktoren möglich. Im Wahlwerbungsexperiment könnte beispielsweise sichergestellt werden, dass jeder Proband den Wahlkampfsport alleine betrachtet, sodass keinerlei Beeinflussung durch andere Probanden gegeben ist. Auch der Geräuschpegel in dem Raum, der sich auf die Aufmerksamkeit des Probanden und damit evtl. auf den Effekt des Spots auswirken könnte, lässt sich auf ein absolutes Minimum reduzieren. Lassen sich Störfaktoren nicht komplett ausschalten, so ist es doch oftmals möglich, zumindest ihren Einfluss durch *Konstanthaltung* zu kontrollieren. Die Logik dahinter ist einfach: Selbst wenn man davon ausgeht, dass beispielsweise die Art und Weise wie der Versuchsleiter den Probanden gegenüber auftritt (im Extremfall gibt er sich beispielsweise als SPD-Mitglied zu erkennen) Auswirkungen auf die vom Probanden geäußerte Wahlabsicht zeitigt, kann man doch dadurch, dass alle Personen vom selben Versuchsleiter angesprochen werden, einer systematischen Verzerrung der Ergebnisse entgegenwirken. Ähnlich ließe sich mit situationsbedingten Störfaktoren umgehen. So würde man allein schon deshalb während der Zeit, in der die Experimentalgruppe den Wahlkampfsport sieht, der Kontrollgruppe einen unverfänglichen Werbefilm vorführen, um dieser eine ansonsten ähnliche Reizumgebung zu präsentieren. Um zusätzlich noch den Effekt auszuschalten, dass der Versuchsleiter (vielleicht sogar unbewusst) sich einzelnen Probanden gegenüber anders verhält, könnte man eine Tonaufzeichnung mit den Instruktionen erstellen und diese für jeden Probanden abspielen, so dass der Einfluss des Versuchsleiters auf ein Minimum reduziert würde und der verbliebene Einfluss (über Intonation, Wortwahl usw.) für alle Probanden konstant wäre. Eine weitere Option wären sogenannte *Blind-* oder *Doppelblindversuche*. Bei ersteren wissen die Probanden nicht ob sie der Treatment- oder der Kontrollgruppe zugeordnet werden oder sogar nichts über die eigentlich im Experiment zu testende Fragestellung. Bei einem Doppelblindversuch wissen darüber hinaus auch der Versuchsleiter, bzw. derjenige, der die Ergebnisse auswertet, nicht über die zugrunde liegende Hypothese Bescheid, so dass sie nicht (unbewusst) einer der Gruppen verzerrende Signale senden, bzw. einen Bias bei der Auswertung erzeugen können (Diekmann 2007, S. 338).

⁷ Der Gegensatz zu diesen sind Feldexperimente, die in der natürlichen Situation des zu untersuchenden Phänomens durchgeführt werden. Genauer zu den beiden Optionen im nächsten Abschnitt.

2.5 Labor- und Feldexperimente

Eine herausragende Rolle für die Art und Weise der Durchführung aber auch für die möglichen Schlüsse, die sich aus einem Experiment ziehen lassen, spielt die Frage, unter welchen Bedingungen es durchgeführt wird: Maximal kontrolliert im Labor oder näher an der lebensweltlichen Realität der Untersuchungsobjekte? Laborexperimente sind darauf ausgelegt, Störfaktoren möglichst umfassend zu kontrollieren. Oftmals generiert die künstliche und für die Probanden ungewohnte, auf jeden Fall nicht ihrem Lebensalltag entsprechende Untersuchungssituation im Labor aber selbst einen nicht zu vernachlässigenden Bias. Die Probanden verhalten sich im Labor anders als in der Realität. Damit mag zwar die interne Validität eines Laborexperiments gegeben sein. Wenn sich dessen Ergebnisse jedoch nicht auf einen Kontext außerhalb des Labors übertragen lassen – es also an der externen Validität mangelt – ist nur wenig gewonnen (vgl. Infobox).

Interne und externe Validität

Ein Experiment hat eine hohe interne Validität, wenn sämtliche potentiellen Störfaktoren kontrolliert werden können und damit eine Veränderung der AV ausschließlich auf das Treatment der UV zurückzuführen ist (McDermott 2002, S. 35). Kann dieser kausale Schluss nicht eindeutig gezogen werden, weil beispielsweise einzelne Drittfaktoren unbekannt sind und damit nicht für die Parallelisierung verwendet werden können oder die Einteilung in Experimental- und Kontrollgruppe nicht randomisiert erfolgen kann, dann reduziert dies die interne Validität. Weitere Gefahren für die interne Validität wären unter anderem gegeben wenn Versuchspersonen sich selbst für ein Experiment melden (Selbstselektion), sich während der Durchführung des Experiments bestimmte Bedingungen ändern bzw. sich für einzelne Probanden unterscheiden, ohne dass der Versuchsleiter dies beeinflussen könnte (wenn z. B. ein Feueralarm ertönt und damit die Konzentration der Versuchsteilnehmer unterbricht) oder das Ergebnis durch Erfahrungen, die während des Experiments erworben werden beeinflusst wird. Ein solcher Lerneffekt wäre beispielsweise gegeben, wenn bei einem Experiment der Pre-Test exakt dieselben Fragen enthält wie der Post-Test. Allein durch die Beschäftigung mit dem Pre-Test könnten sich im Post-Test bereits die Antworten ändern, so dass eine Variation zwischen den beiden nicht automatisch alleine auf die dazwischen während des Experiments erfolgte Manipulation der UV zurückgeführt werden kann (McDermott 2002, S. 36–37).

Von externer Validität spricht man, wenn die Ergebnisse des Experiments generalisierbar sind. Es lassen sich drei Arten der Übertragbarkeit unter-

Methodologie, Methoden, Forschungsdesign
Ein Lehrbuch für fortgeschrittene Studierende der
Politikwissenschaft

Hildebrandt, A.; Jäckle, S.; Wolf, F.; Heindl, A.

2015, XI, 387 S., Softcover

ISBN: 978-3-531-18256-8