



Leseprobe aus Schumann, Experimente contra Kriminalität,
ISBN 978-3-7799-6375-2 © 2021 Beltz Juventa
in der Verlagsgruppe Beltz, Weinheim Basel
[http://www.beltz.de/de/nc/verlagsgruppe-beltz/
gesamtprogramm.html?isbn=978-3-7799-6375-2](http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-7799-6375-2)

Inhalt

Worum geht es? 9

- 1 Goldstandard 9
- 2 Wieso ist Losen wichtig? 12
- 3 14 spannende Experimente 14

Klassiker 17

- 1 Beginn einer Ära – die Cambridge-Somerville Youth Study 17
 - Die Idee* 17
 - An die Arbeit!* 19
 - Was heißt ,betreuen?* 21
 - Spielball der Weltpolitik* 23
 - Und die Kontroll-Jungen?* 25
 - Die Bilanz* 25
 - Ten Years After* 27
 - Stimmen McCords Erklärungen?* 28
- 2 Der Leuchtturm – das Minneapolis Domestic Violence Experiment 31
 - Ein Paukenschlag wird vorbereitet* 31
 - Ein bürokratisches Losverfahren* 33
 - Probleme* 34
 - Wirkungen* 35
 - Unglaublicher Einfluss* 37
- 3 Besser nichts tun? 39
 - Test einer Theorie* 39
 - Spielt die Polizei mit?* 41
 - Labeling fördert Rückfall* 42
 - Die Erklärung* 43
- 4 Zeitquoten statt Los – das Court-Employment-Experiment 45
 - Nötigung* 45
 - CEPs Ansatz* 45
 - Lauter gute Argumente* 47
 - Zeitquoten-Randomisierung* 48

Kleine Mängel 50

Fehlanzeige 52

Abkehr 53

Kontrollierte Tests neuer Maßnahmen 55

5 Sozialtherapie im Gefängnis – ein deutscher Solitär 55

Auf Messers Schneide 55

Für und Wider 57

Gab's das nicht schon mal? 58

Der Entscheid 60

Das Experiment 61

Zäher Zulauf 62

Wie könnte Sozialtherapie denn helfen? 63

Ernüchterung 65

Streit 66

Das letzte Wort 68

6 Brennpunkte entschärfen 70

Plätze statt Personen 70

„broken windows“ 72

Paare von „hot spots“ 73

Das Lowell-Experiment 74

Effektiv? 76

Bewiesen 77

7 Reintegration nach Beschämung – die RISE-Experimente 79

Prävention durch Scham 79

Experimentator gesucht 80

Vier RISE-Experimente 81

Alkohol am Steuer 84

Volatile Wirkungen 86

Powerplay 87

Gleiche Experimente – verschiedene Orte 91

8 Ein gigantisches Experiment mit Strafaussetzung
zur Bewährung 91

Rekordverdächtig 91

-
- Was soll „intensive Bewährungshilfe“ bewirken?* 92
Quer durch die USA 94
Ein Richter gegen Texas 96
Mammutjob in Etappen 98
Guter Start in Kalifornien
(*Contra Costa, Ventura, Los Angeles*) 99
Verkehrte Welt (Marion County, Oregon) 100
Urinproben bei Drogentätern in acht Counties 102
Richterliche Sabotage (Milwaukee, Wisconsin) 104
Entkommt Texas der Strafzahlung? (Dallas, Houston) 105
Kein Erfolg, nirgendwo? 107
Punitiver Zugewinn 109
- 9 Minneapolis fünffach wiederholt 111
 Warum Replikationen? 111
 Das Spouse Assault Replication Program (SARP) 112
 Omaha geht voran 113
 Charlotte (North Carolina) folgt 115
 Larry Sherman ist enttäuscht 118
 Florida gibt Entwarnung 120
 Colorado Springs 122
 Welches Fazit ist richtig? 123
 Letzte oder vorletzte Worte? 124
- 10 Nationale Evaluation von Diversion 127
 Endlich Klarheit schaffen! 127
 Das Memphis-Metro-Experiment 129
 Diversion in New York City 132
 Das Kansas-City-Experiment 134
 Diversion in Orange County, Florida 135
 Triumph der Nullhypothese 137
 Ungläubigkeit 138
- Experimentatoren geraten in Konflikte** 141
- 11 Vom Staatsanwalt bekämpft 141
 Premiere 141
 Kooperationspartner Bewährungshilfe 143

<i>Der Konflikt wird öffentlich</i>	144
<i>Alleingelassen?</i>	146
<i>Hat sich der Stress gelohnt?</i>	148
<i>Hintergründe und Parallelen</i>	150
12 Aus Angst konform?	152
<i>Die Idee</i>	152
<i>Widrigkeiten</i>	154
<i>Kontras</i>	155
<i>Showdown</i>	157
13 Zürcher Gewaltprävention mit Nachspiel	158
<i>Qualitätsansprüche</i>	158
<i>Was bremst Jugendgewalt?</i>	159
<i>Wie funktioniert Triple P?</i>	160
<i>Was will PFADE erreichen?</i>	161
<i>Schulen zu verlosen!</i>	162
<i>Das Schuldepartement handelt</i>	163
<i>PFADE ist Lehrersache</i>	164
<i>Wie wird evaluiert?</i>	166
<i>Gibt es positive Wirkungen?</i>	167
<i>Nichts? Nirgendwo?</i>	169
<i>Ein Nachspiel</i>	170
14 John Henry schlägt zu	173
<i>Richter wünschen ein Experiment</i>	173
<i>Drei Varianten der Bewährung</i>	174
<i>John Henry schlägt zu</i>	176
<i>Resultate</i>	177
Schlussfolgerungen	179
<i>Effekte als Mangelware</i>	179
<i>Gültigkeit</i>	181
<i>Methodische Fortschritte</i>	184
<i>Doch nur Bronzestandard?</i>	187
<i>Wie weiter?</i>	188
Dank	191

Worum geht es?

1

Goldstandard

Maßnahmen, die Straftaten verhindern sollen, werden auf ihre Wirksamkeit getestet. Und zwar durch Zufallsexperimente. Adressaten einer Präventionsstrategie vergleicht man mit Personen, denen die Strategie vorenthalten wird. Dazu werden Adressaten¹ und Vergleichspersonen ausgelost. Indem das Los entscheidet, welche Person in die eine oder die andere Gruppe kommt, werden beide Gruppen vergleichbar, ja austauschbar. Beide Gruppen sind äquivalente Stichproben der Gesamtheit. Begehen die Personen in der Gruppe, die eine Präventionsmaßnahme durchlief, später weniger Straftaten, hat sie offenbar gewirkt und rechtfertigt breite Anwendung.

Wir kennen solche Experimente aus der Pharmaforschung. Eine Gruppe von Patienten bekommt ein neues Medikament, die andere Gruppe ein Placebo. Wer in welche Gruppe kommt, entscheidet das Los. Gibt es einen Wirkungsunterschied, ist das Medikament effektiv, falls nicht, ist es nutzlos. Eindeutig beweisen kann man Wirksamkeit nur so, nur wenn die beiden Patientengruppen vergleichbar sind. Die Bildung der Gruppen durch ein Zufallsverfahren gewährleistet das. Dadurch werden nach der Wahrscheinlichkeitstheorie beide Gruppen gleich gute Stichproben aus dem Universum aller Patienten mit dem spezifischen Leiden.

Genauso sind Experimente mit Maßnahmen gegen Kriminalität gestaltet. Sie werden seit Mitte des 20. Jahrhunderts vor allem in den USA praktiziert, um Fragen zu beantworten wie: Senken bestimmte Therapien im Gefängnis die Gefahr von Rückfälligkeit? Schafft stärkere Polizeipräsenz eine größere Sicherheit für Bürger? Kann für neue Strafformen oder Betreuungsangebote Wirksamkeit, d.h. eine geringere Rückfallrate, nachgewiesen werden?

1 In diesem Buch werden aus Gründen der besseren Lesbarkeit Personen durchweg im generischen Maskulinum benannt (wie hier z. B. Adressaten). Darin sind Männer und Frauen zugleich einbezogen. Wenn Sachverhalte sich auf geschlechtsspezifische Teilgruppen beziehen, also eine Sexusneutralität nicht besteht, wird dies natürlich sprachlich berücksichtigt.

Dass Maßnahmen der Kriminalitätskontrolle experimentell überprüft werden, ist Folge einer negativen Bilanz, die Robert Martinson 1974 über Behandlungsmaßnahmen im Strafvollzug veröffentlichte. Unter der Frage „What works?“ sichtete er alle auffindbaren Evaluationsstudien über Bildungsmaßnahmen, Gesprächstherapien und andere Therapien im Gefängnis. Seine Bilanz war: „Nothing works!“ Überall bleibe die Rückfälligkeit die gleiche, ob jemand eine Therapie bekomme oder nicht.

Martinsons Bilanz erregte Aufruhr und Widerspruch. Die 1970er Jahre, in denen er diese Hiobsbotschaft verkündete, waren ja die Blütezeit des Resozialisierungsgedankens im Strafvollzug. Da schockierte sein: „Nothing works!“. Alle Hilfsangebote für Strafgefangene konnten doch nicht zwecklos sein. Martinson wurde auch deshalb kritisiert, weil die Studien, die er auswertete, methodische Mängel hatten. Man sagte: Ein so fundamentaler Verriss setzt Studien voraus, die methodisch einwandfrei sind. Das sind nur Zufallsexperimente. Nur bei diesen sei die Wirkung oder Wirkungslosigkeit einer Therapie schlüssig bestimmbar.

Konsequent forderte im Jahr 1979 ein Panel der amerikanischen National Academy of Science zu „The Rehabilitation of Criminal Offenders“, dass dringend randomisierte Experimente zur Prüfung der Frage „What works?“ durchgeführt werden müssten. Denn wenn sich herausstellt, dass weniger eingriffsintensive Sanktionen die gleiche Rückfallrate aufweisen, verlangten die Menschenrechte, dass auf die schärfere Maßnahme *verzichtet* werden muss.

Nun wurden und werden immer wieder ethische und rechtliche Bedenken gegen Zufallsexperimente auf dem Gebiet der Strafverfolgung geäußert. Dass Menschen – je nach Losglück – mehr oder weniger stark bestraft werden sollen, erscheint als Willkür, als Verstoß gegen Rechtstaatlichkeit. Deshalb setzte das Federal Judicial Center der USA 1981 Bedingungen fest, die für ein Losverfahren bei Justizentscheidungen gegeben sein müssen. Maßgebliches Kriterium ist, dass die erhoffte positive Wirkung einer neuen Maßnahme *noch unbewiesen* ist. Dann ist unklar, ob sie der Experimentalgruppe nutzt oder gar schadet. Insofern wird die Kontrollgruppe nicht benachteiligt.

Das Justizministerium begann in den 1980er Jahren über seine Forschungsabteilung, das National Institute of Justice (NIJ), explizit für Experimente zu werben, und stellte gleichzeitig erhebliche Forschungsmittel dafür zur Verfügung. Seither gehören Feldexperimente mit Zufallsauswahl zum Alltag kriminologischer Forschung in den USA. Dafür sorgte auch eine Expertise, die 1998 ein Team der University of Maryland um Lawrence W. Sherman für den U. S.

Congress erarbeitete. Unter dem Titel „*Prevention of crime: What works, what doesn't, what's promising?*“ wurden die existierenden Evaluationen von Präventivmaßnahmen auf Schlüssigkeit ihrer Ergebnisse bewertet. Gütekriterium war die Gültigkeit des Forschungsdesigns. Die größte interne Validität, d.h. Eindeutigkeit des Kausahnachweises, wurde Studien zugesprochen, die ein experimentelles Design verwendet hatten. Das setzte Maßstäbe. Das Zufallsexperiment wurde zum *Goldstandard* der Evaluationsforschung erklärt.

Es fand ein Umdenken in der Förderung von Forschung statt, ein sogenannter *experimental turn*: Experimente wurden nachgefragt. Lawrence Sherman gründete zur Popularisierung die *Academy of Experimental Criminology* in der *American Society of Criminology*. Und der Nestor von Experimenten in den Sozialwissenschaften, Robert Boruch, rief im Februar 2000 die *Campbell Collaboration* ins Leben. Nach dem Muster der in der Medizin tätigen *Cochrane Collaboration* wird durch Sichtung des gesamten Forschungsstandes geprüft, welche Kriminalität verhindernden Maßnahmen nachweislich wirksam und somit empfehlenswert sind.² Dabei werden Studien bevorzugt, die als randomisierte Feldexperimente durchgeführt wurden.

Zufallsexperimente mit Maßnahmen in der Polizeiarbeit, in Gerichten oder im Strafvollzug werden ganz überwiegend in den USA durchgeführt. In europäischen Ländern, besonders in Deutschland, haben sie Seltenheitswert. Ethische und rechtliche Bedenken stehen hier im Wege. Das können aber auch vorgeschoßene Argumente sein. Möglicherweise sind in diesen Ländern Polizei, Justiz und Strafvollzug an einer exakten Evaluation ihrer Praxis gar nicht interessiert. Wenn schon Evaluation, dann bitte mit Methoden, deren Ergebnisse man leicht in Zweifel ziehen kann. Doch kann man sich auf Dauer der Logik verschließen, dass präventive Maßnahmen evidenzbasiert sein sollten?³ Neuerdings wächst allerdings das Interesse an Experimenten unter den Sozialwissenschaftlern Deutschlands⁴.

Um die Debatte weiter zu beleben, erzähle ich in diesem Buch 14 interessante Feldexperimente nach, mit denen Maßnahmen gegen Kriminalität getestet werden. Ich verwerte dabei Veröffentlichungen und unveröffentlichte For-

2 <http://campbell.gse.upenn.edu>

3 Dafür werben auch M. Walsh, u.a. (Hrsg.), *Evidenzorientierte Kriminalprävention in Deutschland*, Wiesbaden 2018

4 Vgl. das Sonderheft der Zeitschrift *Soziale Welt*: „Experimente in den Sozialwissenschaften“, 2015

schungsberichte sowie Interviews, die ich mit den meisten Experimentatoren geführt habe. Die Geschichten zeigen die Vor- und Nachteile solcher Designs. Sie stellen Aufwand und Ergebnisse gegenüber. Sie erlauben ein erstes Urteil über Fragestellungen, die für Experimente, die mit Losverfahren arbeiten, geeignet und weniger geeignet sein können.

2

Wieso ist Losen wichtig?

Die Zufallszuweisung (*randomization*) kontrolliert nach der Wahrscheinlichkeitstheorie *alle* relevanten Faktoren. Wird dagegen eine Kontrollgruppe nicht per Los, sondern durch gezielte Auswahl zusammengestellt, kann man nicht sicher sein, ob später gefundene Unterschiede zur Experimentalgruppe auf die Maßnahme oder auf die Selektion zurückgehen. Werden die Gruppen aber durch Los gebildet, ist dieses Problem interner Gültigkeit behoben. Unterschiede beruhen allein auf der erlebten (oder vorenthaltenen) Maßnahme, nicht auf anderen Faktoren. Diese gelten als kontrolliert. Und fast wichtiger: diese Kontrolle ist auch für alle Faktoren gegeben, *deren Messung unmöglich* ist, oder *die unbekannt* sind, oder deren Relevanz sich erst nach Ende der Forschung herausstellt⁵.

Randomisierung ist aber oft nicht möglich. Richter oder Behörden bestehen darauf, die Maßnahmen mit Blick auf den Einzelfall zuzuweisen und lassen nicht zu, dass Forscher die zu untersuchende Sanktion oder deren Alternative bei Probanden je nach Los verteilen. Die von Richtern bevorzugte gezielte Auswahl von Personen für eine Reaktionsweise schränkt aber die Gültigkeit von Evaluationsergebnissen ein. Diejenigen, denen die Maßnahme vorenthalten wurde, müssen ja, gerade weil sie als ungeeignet für die zu evaluierende Maßnahme galten, von den Ausgewählten verschieden sein.

Ein Design mit einer gezielt ausgesuchten Kontrollgruppe wird Quasi-Experiment genannt. Hier versucht der Forscher, eine Kontrollgruppe zusammenzustellen, die nach allen relevant erscheinenden Merkmalen der Experimentalgruppe ähnelt. Aber Unterschiede bleiben. Vor allem kann man durch diese gezielte Auswahl natürlich unbekannte Drittvariablen, die vielleicht die Wir-

5 Die beste Einführung in das Thema bietet Robert F. Boruch, *Randomized Experiments for Planning and Evaluation: A Practical Guide*, Thousand Oaks 1997.

kung hervorgerufen haben, nicht kontrollieren. Im Zufallsexperiment darf man sie aufgrund der Wahrscheinlichkeitstheorie als in Experimental- wie Kontrollgruppe ähnlich verteilt unterstellen. Quasi-Experimente laden daher dazu ein, an der Trifigkeit der Ergebnisse zu zweifeln. Gleichwohl sind Institutionen eher bereit, ein Vorgehen zu akzeptieren, das die Gruppen gezielt und nicht per Zufall auswählt. Umso anspruchsvoller ist es, ein Randomexperiment durchzuführen.

Und viel komplizierter. Das macht schon der Vergleich mit den bei Psychologen beliebten Laborexperimenten deutlich. Im Labor kann der Experimentator schalten und walten, wie es ihm beliebt. Die Versuchspersonen, ob in die Experimental- oder Kontrollgruppe gelost, machen alles mit, was der Versuchsleiter von ihnen verlangt. Die berüchtigten Experimente von Milgram, der Versuchspersonen dazu verleitete, starke Stromstöße als Sanktionen für Fehlleistungen anderer auszuteilen, haben dies gezeigt. Die Versuchspersonen werden oft bezahlt oder sind Studenten, die durch Studienbedingungen zur Teilnahme an Tests verpflichtet sind. Laborexperimente sind Spielplätze, Feldexperimente sind das wahre Leben.

Beim Feldexperiment greift der Forscher gewissermaßen als Herrscher auf Zeit in einen sozialen Raum ein. Zunächst muss er eine Institution davon überzeugen, dass entweder deren eingespielte Praxis oder eine neue Methode auf Effektivität geprüft werden soll. Wenn sie zustimmt, muss die Kooperation der Mitarbeiter erreicht werden; diese machen allerdings nicht immer widerstandslos alles, was die Spitze der Institution will. Sie haben keine Lust, ihre täglichen Routinen zu ändern. Im Feldexperiment werden entweder sie oder ihre Klientel Adressaten eines Losverfahrens. Diese Zufallsauswahl will der Forscher selbst vornehmen, wodurch er in die Autonomie der Praktiker eingreift. In ihrer täglichen Arbeit können sie aber die Vorgaben des Experimentators leicht sabotieren. Insbesondere dann, wenn sie an Ergebnissen des Evaluation, z. B. an Neuerungen, nicht interessiert sind. Dies bewirkt die Störungsanfälligkeit von Randomexperimenten in sozialen Räumen. Beispiele dafür gibt es zu Hauf. Die hier erzählten Fallgeschichten bieten denn auch ein Kaleidoskop möglicher Stolpersteine, Hindernisse und Fehlschläge solcher Feldexperimente.

3

14 spannende Experimente

Schon vor ihrem Boom seit den 1980er Jahren haben randomisierte Feldexperimente in der amerikanischen Kriminologie eine gewisse Tradition. Es gibt einige markante Beispiele. Das bekannteste ist die Cambridge-Somerville Youth Study zur Verhinderung von *Jugendkriminalität*, die ich in Kapitel 1 erzähle. Ferner der Meilenstein der Evaluation von *Polizeieingriffen bei häuslicher Gewalt*: das Minneapolis Domestic Violence Experiment von Lawrence Sherman. (Kapitel 2). Dieses Experiment hat den Durchbruch der Zufallsexperimente im Feld sozialer Kontrolle bewirkt. Das in Kapitel 3 skizzierte Experiment testet eine wichtige kriminologische Theorie, die *Etikettierungstheorie* (labeling theory), nach der Strafen durch ihr Stigma Rückfälle fördern können. Kapitel 4 beschreibt ein Experiment in der Justiz, das zu scheitern drohte aufgrund ethischer und rechtlicher Bedenken von Strafverteidigern gegen die Auslösung ihrer Mandanten für Anklageverzicht. Erst der geniale Einfall einer *zeitabhängigen Zufallszuweisung* macht die Durchführung möglich. Diese vier Experimente können als *Klassiker* gelten.

Der zweite Teil des Buches ist Experimenten mit *neuen Maßnahmen* gewidmet. Kapitel 5 erzählt von dem einzigen Zufallsexperiment, das bislang in Deutschlands Strafvollzug genehmigt wurde, einem Solitär also, mit dem der Nutzen der neuen Maßnahme „*Sozialtherapie*“ überprüft worden ist. Wie die grundsätzlichen Widerstände und rechtlichen Gegenargumente ausgeräumt werden, ist instruktiv. In Kapitel 6 werden nicht Personen, sondern Orte, „*hot spots*“, Brennpunkte des Kriminalitätsgeschehens, für unterschiedliche Polizeiinterventionen ausgelost. An einigen ist die Polizei stark präsent, an anderen mit den üblichen Streifenfahrten. Es geht um die Kernfrage der Polizeiarbeit: Zeigt Präsenz Wirkung? Kapitel 7 stellt Randomexperimente dar, mit denen in Australien eine neue Sanktion, nämlich *Beschämung von Straftätern* durch Konfrontation mit Opfern, getestet wurde. Später wurde sie per Gesetz eingeführt.

Teil 3 befasst sich mit Experimenten, die *gleichzeitig an mehreren Orten* durchgeführt werden. Der Gedanke ist aus den Naturwissenschaften übernommen. Experimente müssen mehrfach durchgeführt werden, parallel und wiederholt, ehe man ihre Resultate als gültig betrachten kann. Vielleicht sind in einer einzelnen Experimentsituation unerkannte spezielle Rahmenbedingungen im Spiel und bewirken die positiven Effekte? Replikationen wollen die sogenannte