
Ambulante Sanktionen im Vergleich der experimentellen Evaluationsforschung

C

I **Vergleich der Wirkungsweise verschiedener Sanktionen und das Problem der Kausalität**

Zusätzlich zu der Problematik des Vergleichs von Sanktionen zwischen unterschiedlichen kulturellen und rechtlichen Kontexten sind Vergleiche bezogen auf Sanktionen auch noch in anderer methodologischer Hinsicht ausgesprochen problematisch. Wenn es um die empirische Erforschung der Wirkung von unterschiedlichen Sanktionen im Vergleich zueinander geht, etwa in Hinblick auf ihre Effektivität zur Prävention von Rückfällen, stellt sich die Frage nach der methodischen Konzeption von Studien, mit denen solche möglichen Einflüsse auf die Rückfallwahrscheinlichkeit überhaupt feststellbar werden. Dieser Frage soll im Folgenden näher nachgegangen werden. Dabei ist die Reduktion von Rückfallquoten zwar nur ein mögliches Kriterium für den Erfolg von strafrechtlichen Sanktionen, jedoch ist sie die Variable, welche von der Forschung am häufigsten herangezogen wird. Obwohl es sehr gute Argumente dafür gibt, dass außer der Rückfallquote auch andere Wirkungen von Sanktionen berücksichtigt werden sollten, etwa solche in Hinblick auf den jeweils mit ihnen verbundenen Grad der Integration oder Desintegration der Betroffenen in die Gesellschaft, ist die Fokussierung auf das Rückfallkriterium nicht auf bloße Engstirnigkeit der Forschenden zurückzuführen, sondern erscheint aus dem Blickwinkel des Sanktionenrechts als nachvollziehbar. Wenn es für einen Staat nämlich zulässig sein soll, in die Freiheit oder andere Grundrechte von Betroffenen einzugreifen, dann muss er

stichhaltige Rechtfertigungsgründe für diese Eingriffe haben. Davon könnte allenfalls dann gesprochen werden, wenn Grund zu der Annahme bestünde, dass die Sanktion einen legitimen Zweck verfolgte, dessen Erreichen auch realistisch wäre. Hingegen verstößt die Verhängung von Sanktionen entgegen der Erkenntnis, dass diese im Hinblick auf das staatlich verfolgte Ziel wenig erfolgsversprechend sind, gegen den Grundsatz der Verhältnismäßigkeit. Daher ist es wichtig, die Wirksamkeit von Sanktionen (auch) in Bezug auf ihren gesetzlich festgelegten Zweck, die Rückfälligkeit, zu bewerten.

Die Nummern 89 und 90 der Europäischen Grundsätze für nicht-freiheitsentziehende Strafen und Maßnahmen verlangen von den Staaten Anstrengungen, den Beitrag ambulanter Sanktionen zur Verminderung von Gefangenzahlen und in Bezug auf die Berücksichtigung der Bedürfnisse von Straffälligen soweit diese sinnvollerweise mit dem Ziel der Rückfallprävention in Zusammenhang gebracht werden können, zu evaluieren. Weiterhin soll nach den Grundsätzen die Effizienz in Hinblick auf die dabei aufgewendeten finanziellen Mittel sowie der Beitrag ambulanter Sanktionen zur Kriminalitätsreduktion evaluiert werden. Hervorzuheben ist, dass ausdrücklich Forschung zu der Frage verlangt wird, ob ambulante Sanktionen tatsächlich als Alternative zu freiheitsentziehenden eingesetzt werden oder ob sie lediglich sonstige Interventionsformen, etwa eine Geldstrafe, ersetzen.⁷² Im Jahr 2000 veröffentlichte das Ministerkomitee Empfehlungen an die Mitgliedstaaten zur Verbesserung der Umsetzung der Europäischen Grundsätze über die nicht-freiheitsentziehenden Strafen und Maßnahmen. Nummer 24 dieser Leitlinien für eine breitere und effektivere Anwendung ambulanter Sanktionen und Maßnahmen empfiehlt u. a. eine Wirkungsanalyse, Nummer 27 schlägt darüber hinaus einen Vergleich verschiedener Ansätze und von deren Wirksamkeit vor, und Nummer 26 weist auf die Notwendigkeit hin, die Perspektive der von ambulanten Sanktionen und Maßnahmen Betroffenen auf diese zu untersuchen. Während die wichtige Frage nach der Betroffenenperspektive bereits oben angesprochen worden ist, ist die Auseinandersetzung mit Fragen der Wirksamkeit verschiedener Sanktionen (outcome evaluation), auch im Vergleich zueinander, Gegenstand dieses Abschnittes. In Bezug darauf empfiehlt das Ministerkomitee eine gründliche empirische Untersuchung mit sog. rigorosen Forschungsmethoden für eine auf wissenschaftliche Erkenntnisse gestützte Anwendung von Sanktionen und Maßnahmen. Experimentelle Studien mit einer randomisierten Zuweisung von Personen zu Ver-

72 Vgl zu Nr. 89 und 90 der Grundsätze auch die offiziellen Erläuterungen: Explanatory Memorandum to Recommendation No. R (92) 16 of the Committee of Ministers to member states <https://wcd.coe.int/ViewDoc.jsp?id=615649> (Zugegriffen: 20. März 2015)

gleichs- und Kontrollgruppen und quasi-experimentelle Studien werden in diesem Zusammenhang als aussagekräftige Forschungsdesigns betrachtet. Solche experimentellen Studien sollen dem Ministerkomitee zufolge mit gebührender Berücksichtigung ethischer Erwägungen durchgeführt werden. Allerdings findet sich an keiner Stelle eine Erläuterung oder ein weiterführender Hinweis wie die Frage beantwortet werden soll, ob und ggf. wie entsprechende Forschungsvorhaben in Einklang mit rechtlichen und ethischen Anforderungen im Bereich des strafrechtlichen Sanktionsrechts gebracht werden können. Diese Frage ist nämlich weithin ungeklärt und insoweit ist jedenfalls die randomisierte Zuweisung von Personen zu unterschiedlichen Sanktionsformen oder sogar zu Sanktionen im Vergleich zum Verzicht auf diese aus juristischer Sicht kritisch zu betrachten.⁷³ Möglicherweise auch aufgrund dieser Probleme und den hohen Kosten experimenteller Studien sind – trotz eines Anstiegs in den vergangenen Jahrzehnten – immer noch nur sehr wenige experimentelle Studien zu diesem Thema zu verzeichnen. Das gilt besonders für Deutschland und andere nicht-englischsprachige Länder in Europa. Ein weiterer Grund für die geringe Anzahl experimenteller Studien mag die Skepsis einiger, insbesondere qualitativ arbeitender Forscher und Forscherinnen gegenüber der methodologischen Überlegenheit solcher Studien sein, die von anderen jedoch mit weitgehendem Absolutheitsanspruch und wachsendem Einfluss hervorgehoben wird. Viel lässt sich zur durchaus vorhandenen Berechtigung solcher und weiterer Einwände sagen, wenn auch nicht an dieser Stelle.⁷⁴

Im Bereich der internationalen (westlichen) wissenschaftlichen Gemeinschaft der quantitativen kriminologischen Forschung wurde zur Jahrtausendwende die Campbell Collaboration gegründet, welche sich zum Ziel gesetzt hat, u. a. im Bereich der Strafjustiz experimentelle Forschung zu fördern und die aus ihr gewonnenen Ergebnisse zu analysieren.⁷⁵ Mit systematischen Forschungsübersichten (Systematic Reviews) – einschließlich Metaanalysen – zu bisherigen veröffentlichten und unveröffentlichten experimentellen und quasi-experimentellen Studien und der Umsetzung ihrer Ergebnisse in Politikentscheidungen wird eine rationale Kriminalpolitik angestrebt. Im Jahr 2006 veröffentlichten Villettaz et al. eine solche systematische Forschungsübersicht der Campbell Collaboration über die Auswirkung von freiheitsentziehenden im Vergleich zu ambulanten Sanktionen auf die Rückfälligkeit von Straffälligen.⁷⁶ Obwohl ihnen eine überwältigende Anzahl von Studien (mehr als 3000 Abstracts) über Rückfälligkeit nach Einwirkung

73 Vgl. Graebisch 2000, S. 271-282; dies. 2007 S. 193-204.

74 Vgl. zum Ganzen Graebisch 2009; 2014 b.

75 www.campbellcollaboration.org.

76 Villettaz et al. 2006.

unterschiedlicher Sanktionen vorlagen, erfüllten nur 23 von ihnen die Kriterien für Aufnahme in die Campbell-Review, wobei wiederum nur fünf dieser Studien auf einem randomisierten oder quasi-experimentellen Forschungsdesign basierten. Insgesamt konnten 27 Vergleiche zwischen der Rückfallquote von Verurteilten mit Freiheitsstrafen einerseits und Verurteilten mit ambulanten Sanktionen andererseits, herangezogen werden. 13 von diesen Vergleichen lieferten signifikante Ergebnisse, wobei elf eine geringere Rückfallquote bei ambulanten Sanktionen und zwei bei Freiheitsstrafen aufwiesen. 14 der 27 zugrunde gelegten Vergleiche ergaben keinerlei statistisch signifikante Differenzen bezüglich der Rückfälligkeit. Villettaz et al. führten zusätzlich eine Metaanalyse auf der Basis der vier randomisierten Studien und einem natürlichen Experiment durch. In dieser kommen die Autoren zu dem Ergebnis, dass ambulante Sanktionen – von Zufallseffekten abgesehen – keine Vorteile, aber auch keine Nachteile bezüglich der Rückfallquoten aufweisen. Gegenteilige Forschungsergebnisse anderer (zugunsten ambulanter Sanktionen) seien ihrer Ansicht nach wahrscheinlich durch Unterschiede zwischen der Gruppe von zu einer Freiheitsstrafe Verurteilten und der Gruppe derjenigen, gegen die eine ambulante Sanktion angeordnet wurde, bedingt. Diesen anderen Studien fehle es an einer methodologischen Kontrolle solcher vorher bestehender Unterschiede.⁷⁷

Wie die andere, unter der Überschrift „evidence-based crime prevention“ verstärkt experimentelle Forschung einfordernde Literatur⁷⁸, so verweist auch diese Systematic Review auf ein Problem, welches sich anhand der seit 2003 im Auftrag des deutschen Bundesjustizministeriums veröffentlichten Statistik zur Rückfälligkeit von auf unterschiedliche Art strafrechtlich Sanktionierten veranschaulichen lässt.⁷⁹ Mittels der Rückfallstatistik können strafrechtliche Folgeentscheidungen gegen Personen miteinander verglichen werden, gegen die vorher einer Freiheitsstrafe mit oder eine ohne Bewährung vollstreckt wurde, gegen die eine ambulante Sanktion verhängt oder bei denen die Verhängung einer Jugendstrafe nach dem Jugendgerichtsgesetz ausgesetzt wurde. Es ist auffällig, dass bei diesem Rückfälligkeitsvergleich Personen, die vorher eine freiheitsentziehende Sanktion (Freiheitsstrafe, Jugendstrafe oder Jugendarrest) verbüßt hatten, wesentlich schlechter abschnitten als diejenigen, gegen die zuvor eine ambulante Sanktion angeordnet worden war. Dies könnte zu der Schlussfolgerung verleiten, dass freiheitsentzie-

77 Villettaz et al. 2006.

78 Vgl. dazu näher (neben den bereits genannten Veröffentlichungen) auch Graebisch 2004, S. 266-283; dies. 2006, S. 46-52; dies. 2011, S. 137-147; dies. 2014, S. 85-100.

79 http://www.bmjv.de/DE/Ministerium/Fachthemen/AbtII/IIA7/Rueckfallstatistik_doc.html (Zugegriffen: 1. März 2015).

hende Strafen einen quantitativen und in Bezug auf die Art der begangenen Straftaten qualitativen Anstieg von Rückfällen bewirkten, die in härteren Strafen bei Folgeentscheidungen münden. So waren nach früherer Verurteilung zu einer Freiheitsstrafe die späteren Verurteilungen sehr viel schwerwiegender als diejenigen, die sich an eine vorangegangene Verurteilung zu einer ambulanten Sanktion anschlossen. Ambulante Sanktionen, so ist diesen Ergebnissen zu entnehmen, erhöhen die Wahrscheinlichkeit von Rückfällen und einer harten späteren Strafsanktion im Vergleich zu freiheitsentziehenden Strafen nicht, im Gegenteil. In diesem Zusammenhang führt die Anordnung von zur Bewährung ausgesetzten Freiheitsstrafen zu einer geringeren Rückfallquote als es bei Haftstrafen ohne Bewährung der Fall ist. Nichtsdestotrotz kann aus den Ergebnissen nicht gefolgert werden, dass ambulante Sanktionen im Hinblick auf die Prävention von Rückfällen erfolgsversprechender wären als die freiheitsentziehenden Sanktionen, da die Gruppen von Betroffenen der jeweiligen Sanktionsart bezüglich verschiedener Variablen, die für die Rückfälligkeit von Belang sein können, nicht vergleichbar sein dürften. Es spricht vielmehr vieles dafür, dass die vorhandenen Sanktionen von den Entscheidungsträgern auf bestimmte Zielgruppen in unterschiedlicher Art und Weise angewandt werden. Das deutsche Strafrecht sieht vor, dass eine Freiheitsstrafe zur Bewährung auszusetzen ist, wenn zu erwarten ist, dass die verurteilte Person künftig auch ohne die Einwirkung des Strafvollzugs keine Straftaten mehr begehen wird oder wenn nach der Gesamtwürdigung von Tat und Persönlichkeit des Verurteilten besondere Umstände vorliegen.⁸⁰ Hieraus kann jedoch weder geschlossen werden, dass die Gerichte sich an diesen Kriterien tatsächlich orientieren noch schließt es das Vorhandensein weiterer Kriterien aus, die für Bewährungsentscheidungen in der Rechtspraxis eine wichtige Rolle spielen.⁸¹ Zudem bringt die besagte Rechtsnorm natürlich auch keinerlei Sicherheit darüber mit sich, dass die seitens des Gerichts angestellten Prognosen über das zukünftige Verhalten der verurteilten Person, insbesondere hinsichtlich der Begehung weiterer Straftaten, auch tatsächlich richtig wären. Umgekehrt wäre es allerdings anmaßend, davon auszugehen, dass die unterschiedlichen Rückfallquoten überhaupt nicht von der Selektion verschiedener Zielgruppen abhängen, die Entscheidungen der Gerichte also in Hinblick auf die prognostischen Erwägungen bei Strafaussetzungen zur Bewährung nicht besser wären als der Zufall. Für einen solchen Schluss liefert

80 Siehe § 56 StGB.

81 Für empirische Erkenntnisse über solche nicht intendierten Einflüsse auf Entscheidungsprozesse von juristischen Entscheidungsträgern mit Bezugnahme auf den Einfluss von Sanktionsempfehlungen, die durch Würfeln gewonnen wurden, English et al. 2006, S. 188-200.

die Forschung nämlich keine Anhaltspunkte. Es ist im Gegenteil sehr naheliegend, wenn auch andererseits wiederum nicht zwingend, dass die Gerichte ihren Entscheidungen Kriterien zugrunde legen, die einen tatsächlichen Einfluss auf die Rückfallwahrscheinlichkeit haben. Ein geringeres Maß an Rückfälligkeit nach der Anwendung ambulanter Sanktionen erlaubt daher nicht den Schluss einer gesteigerten Effektivität gegenüber freiheitsentziehenden Maßnahmen.⁸² Eine solche Schlussfolgerung wäre vielmehr in dem Maße falsch, in dem die zugrundeliegende Gerichtsentscheidung richtig war.

Experimentelle Forschung versucht, diesem Kausalitätsproblem durch die Zufallszuweisung von Personen zu Versuchs- und Kontrollgruppen, welche dann als annähernd vergleichbar gelten, zu begegnen. In diesem Zusammenhang kann es jedoch nur einen allerersten Schritt in Richtung einer interdisziplinären, kulturell sensiblen und kritischen Analyse darstellen, die für ein Verständnis von ambulanten Sanktionen in Europa vonnöten ist,⁸³ wenn die vorhandenen Forschungsergebnisse aus der „What works“-Bewegung berücksichtigt werden. Es ist zudem so, dass diese auch ihrerseits Probleme mit sich bringen, auf die exemplarisch im Folgenden noch näher einzugehen sein wird.

II Randomisierte Kontrollgruppenstudien außerhalb Europas

Eine der vier randomisierten kontrollierten Studien aus der bereits erwähnten Metaanalyse verglich ein intensives Überwachungsprogramm für männliche Jugendliche in Detroit (Michigan) mit der staatlichen Heimeinweisung. Nach einer zweijährigen Nachbeobachtung lag eine gemischte Bilanz vor, das Gesamtfazit ergab jedoch, dass die Versuchsgruppe unter intensiver Überwachung in Bezug auf Rückfälligkeit nicht schlechter abschnitt als die Kontrollgruppe von eingewiesenen Jugendlichen. In der zweiten experimentellen Studie wurden Zweittäter schwerwiegenderer Straftaten (felony) in Oakland County (Michigan) einem Be-

82 Hier ist anzumerken, dass eine solche Annahme noch aus weiteren Gründen fragwürdig erscheint: So darf aus einer eingriffsintensiven Folgeentscheidung nicht automatisch auf eine größere Häufigkeit oder eine gesteigerte Intensität von Straftaten geschlossen werden. Es ist zudem keine Selbstverständlichkeit, dass sich die eine oder andere Sanktionsform als allgemein überlegen erweise, selbst wenn der Vergleich generell auf ambulante versus freiheitsentziehende Maßnahmen beschränkt würde. Tatsächlich besteht vielmehr Anlass zu der Annahme, dass die Effektivität für unterschiedliche Person und Straftaten verschieden ist.

83 McNeill und Beyens 2013, S. 9.

währungsprogramm im Vergleich zu einer Haftstrafe zugewiesen. Während die Gefängnisgruppe eine Rückfälligkeitsquote von 33 Prozent aufwies, fielen bei denjenigen, die dem umfangreichen Bewährungsprogramm zugeteilt wurden, nur 14 Prozent mit einem Rückfall auf. In einer weiteren Studie wurden Jugendliche in Boise (Idaho) einer Gruppe, die eine Wiedergutmachung leisten musste, und im Vergleich dazu einer Gruppe, die eine traditionelle Strafe (Bewährungsstrafe oder Haft) verbüßen musste, zugeteilt. Hierbei wies die Wiedergutmachungsgruppe eine geringere Rückfallquote, sowohl in Hinblick auf die Prävalenz- als auch auf die Inzidenz auf, die Differenzen waren jedoch statistisch nicht signifikant.

III Natürliches Experiment in den Niederlanden: Freiheitsstrafe im Vergleich zur Bewährungsstrafe

Eine weitere Studie, die von Villetta et al. in ihrer Campbell-Review berücksichtigt haben, ist ein – schon älteres – natürliches Experiment aus den Niederlanden.⁸⁴ Anstelle einer Zufallszuweisung nutzten die Forscher dafür eine Situation, die aus königlicher Begnadigung resultierte und verglichen in ihrer Studie die Rückfälligkeit verschiedener verurteilter Personen, die eine nicht zur Bewährung ausgesetzte Haftstrafe von bis zu 14 Tagen erhalten hatten. Aufgrund der königlichen Begnadigung wurde die Strafe von Personen, die ihre Tat vor einem bestimmten Stichtag begangen hatten, ausgesetzt, während die Strafen für Taten, die nach diesem Tag begangen wurden, nicht ausgesetzt wurden. Unterschiede zwischen den beiden Gruppen wurden folglich nur bezüglich des Zeitpunktes der Tat erwartet. Nach dem Ablauf einer Nachbeobachtungszeit von sechs Jahren wiesen die beiden Gruppen ähnliche Rückfallquoten in Bezug auf Verkehrs- und Eigentumsdelikte auf. Im Hinblick auf Gewaltstraftäter wurden jene, deren Strafe aufgrund der königlichen Begnadigung zur Bewährung ausgesetzt worden war, signifikant weniger rückfällig als diejenigen, die eine Haftstrafe verbüßt hatten. Während sich die Haftstrafe bei Gewaltstraftätern also als schädlich erwies, bewirkte die Bewährungsstrafe bei keiner der untersuchten Gruppen negative Effekte hinsichtlich der Rückfälligkeit.

84 van der Werff 1979.

IV Randomisierte Kontrollgruppenstudie in der Schweiz: Gemeinnützige Arbeit im Vergleich zu (sehr) kurzen Freiheitsstrafen

Da Forschungsergebnisse aus den Vereinigten Staaten nur einen begrenzten Wert für den europäischen Kontext haben, kann das Beispiel eines randomisierten Experiments aus der Schweiz vielleicht mehr Aufschluss bringen – ungeachtet der Tatsache, dass die Vergleichbarkeit verschiedener europäischer Staaten und selbst verschiedener Regionen eines Landes ebenfalls problematisch erscheint. Zwischen 1993 und 1995 wurde in dem schweizerischen Kanton Vaud gemeinnützige Arbeit als Alternative für nicht zur Bewährung ausgesetzte kurze Freiheitsstrafen angewandt. Die kurzen Freiheitsstrafen durften nicht länger als 14 Tagen andauern – eine Haftdauer, die in vielen europäischen Ländern wegen der weitverbreiteten Annahme, dass kurzzeitiger Freiheitsentzug schädlich sei, ohnehin unmöglich wäre. Letztere beruht auf der Lehre von Franz von Liszt, der zu Beginn des 20. Jahrhunderts die Überzeugung vertrat, dass kurzfristige Freiheitsstrafen zwar lange genug seien, um den Täter aus seinem Umfeld zu reißen sowie sein Arbeits- und Privatleben zu gefährden, jedoch zu kurz, um eine Gelegenheit zur Besserung zu bieten, die dem Freiheitsentzug jedoch nunmehr als Aufgabe zugeschrieben wurde. In der Schweiz hingegen sind kurzzeitige Freiheitsstrafen sehr beliebt. So betragen die Freiheitsstrafen zum Zeitpunkt des Experiments, d.h. im Jahr 1996, in 42 Prozent der Fälle 14 Tage oder weniger. Zudem wurden in der Schweiz acht Arbeitsstunden als Ersatz für einen Hafttag angerechnet, während die Anzahl an Arbeitsstunden, die etwa aktuell in Deutschland einem Tag in der Haft (Ersatzfreiheitsstrafe) entspricht, normalerweise zwischen drei und sechs Stunden variiert. Die schweizerische Studie ergab, dass die Prävalenz von erneuten Festnahmen durch die Polizei bei der Kontrollgruppe, also denjenigen mit einer Freiheitsstrafe, etwas höher war, jedoch die Unterschiede der beiden Gruppen in Bezug auf Rückfälligkeit, spätere Arbeitsaufnahme sowie die sozialen und privaten Lebensumstände nicht signifikant waren.

Zunächst einmal ist damit festzuhalten, dass die beiden Sanktionsformen keine signifikant unterschiedlichen Ergebnisse erbringen und daher kein Grund erkennbar wird, unter dem Gesichtspunkt der Wirksamkeit betreffend den Rückfall die eingriffsintensivere Sanktionsform zu wählen. Auch wenn in dieser Studie die sehr kurze Freiheitsstrafe keine bessere Wirksamkeit in Bezug auf den Rückfall zeigte, sollten diese Ergebnisse weiterhin nicht zu der Schlussfolgerung verleiten, dass Freiheitsstrafen generell keinen schädlichen Einfluss auf die berufliche Karriere, die soziale Inklusion sowie das Familien- und Privatleben hätten. Während einer (sehr) kurzen Freiheitsstrafe mag eine abhängig beschäftigte Person noch in der

Lage sein, ihr Anstellungsverhältnis aufrecht zu erhalten, verbüßt sie jedoch eine längere Haftstrafe, ist dies gänzlich anders. Hierbei gilt es zu berücksichtigen, dass die Haftdauer in dem vorliegenden Experiment so kurz war, dass sie während des Urlaubs oder gar während des Wochenendes hätte verbüßt werden können. Gemeinnützige Arbeit wurde hierbei zumeist als Alternative zur „Halbgefangenschaft“ angewandt. In der Schweiz gibt der Vollzug der „Halbgefangenschaft“ den Gefangenen die Möglichkeit, die Haftanstalt tagsüber zu verlassen, um arbeiten zu gehen, und verpflichtet sie, lediglich in der Nacht und an Wochenenden in der Haftanstalt zu bleiben. Die Personen aus der Kontrollgruppe der vorliegenden Studie, die also der Inhaftierung zugewiesen wurden, dürften ihre Strafe ohne solche Erleichterungen verbüßt haben, hatten jedoch vor der Zufallszuweisung die Möglichkeit, aus dem Experiment auszusteigen, wovon in der Hoffnung, in den Genuss von Halbgefangenschaft zu kommen, auch Gebrauch gemacht wurde. Dass die Inhaftierung keinen schädlicheren Einfluss auf die spätere Jobsituation genommen hat als die Ableistung von gemeinnütziger Arbeit mag daher weniger bemerkenswert erscheinen, da diejenigen mit einer starken Bindung zu ihrem Job bewusst nicht an dem Versuch teilgenommen haben werden. Es mag ferner sein, dass die negativen Auswirkungen von Freiheitsstrafen auf die Jobaussichten und Lebensumstände der Betroffenen nur in Verbindung mit einer mehr als nur minimalen Haftdauer zum Tragen kommen und dass die jahrhundertelange Debatte über die schädlichen Effekte von kurzen Freiheitsstrafen fälschlicherweise den Blick von den ungewünschten Auswirkungen längerer Haftstrafen abgewendet hat.⁸⁵ Es ist sehr gut denkbar, dass die von der schweizerischen Studie umfasste Haftdauer einen zu kleinen Teil dessen widerspiegelt, was in vielen anderen europäischen Ländern noch als „kurze Freiheitsstrafe“ angesehen wird – wie etwa in Deutschland, wo bis zu sechs Monaten von einer kurzen Freiheitsstrafe ausgegangen wird⁸⁶, um auch nur solche in vergleichbarer Weise abzudecken, von längeren Haftstrafen ganz zu schweigen. Deshalb können die Erkenntnisse aus dieser Studie nicht die Annahme widerlegen, dass kurze – wie auch längere – Haftstrafen zu negativen Effekten bei den Betroffenen führen können.

Ein weiteres Ergebnis der Studie lies den Schluss zu, dass gemeinnützige Arbeit dem Freiheitsentzug vorzuziehen sei. Die zwei Gruppen des Experiments entwickelten sich im Laufe der Nachbeobachtungszeit unterschiedlich. Während die gemeinnützige Arbeit leistende Versuchsgruppe eine Verbesserung bezüglich der Rückfälligkeit aufwies, verschlechterte sich die Kontrollgruppe der Gefangenen gleichzeitig.

85 Für letzteres auch Killias et al. 2000, S. 40-57.

86 § 47 StGB.

Im Hinblick auf die vorliegend interessierende Frage des Vergleichs der Wirkungen von freiheitsentziehenden in Relation zu ambulanten Sanktionen, stellte dieses Experiment – als eine von fünf Studien der besagten Metaanalyse – einen inhaltlich sehr begrenzten Vergleich an, indem eher gleichartige Sanktionen miteinander kontrastiert wurden. So dürfte etwa die Unterschiedlichkeit von einer Freiheits- zu einer Bewährungsstrafe in vielerlei Hinsicht augenfälliger sein als jene zwischen einer Haftstrafe von wenigen Tagen im Vergleich zu wenigen Tagen gemeinnütziger Arbeit. Man würde daher bei der Konzeption einer Studie über die Auswirkungen von Freiheitsstrafen auf die Rückfälligkeit und das Leben von Gefangenen, regelmäßig einen Zeitrahmen wählen, der 14 Tage deutlich übersteigt, will man aussagekräftige Ergebnisse erhalten.

Als ein weiteres Resümee ihrer Studie beschreiben die Autoren, dass beide Sanktionsformen die Straffälligkeit der Betroffenen in Bezug auf die Zahl der von ihnen ausgehenden Taten um 40 Prozent und in Bezug auf den Anteil der Rückfälligen um etwa die Hälfte reduziert hätten. Infolgedessen betrachten die Autoren die Behauptung „punishment does not work“ als ungerechtfertigten Pessimismus. Diese Schlussfolgerung kann allerdings nicht mit dem Untersuchungsdesign und den Ergebnissen der Studie belegt werden und gerade Forscher, die Kausalitätsproblemen mittels experimenteller Forschungsdesigns große Aufmerksamkeit schenken, sollten sich nicht zu derartigen Schlüssen verleiten lassen. Solange experimentelle Designs keine Kontrollgruppen mit Nicht-Intervention einbeziehen, kann nämlich nicht ausgeschlossen werden, dass beide der untersuchten Gruppen aufgrund natürlicher, d.h. nicht auf die jeweilige Intervention zurückzuführender, Reifungsprozesse eine Verbesserung aufweisen. Weil diese dann auch bei bloßem Nichtstun eingetreten wäre, kann nicht auf einen kausalen Einfluss der strafrechtlichen Sanktionen rückgeschlossen werden. Es ist vielmehr sogar denkbar, dass die einbezogenen Personen noch umso bessere Ergebnisse in Hinblick auf spätere Straffälligkeit, ihre berufliche Entwicklung und ihre persönlichen Verhältnisse gezeigt hätten, wenn mit überhaupt keiner Sanktion auf sie eingewirkt worden wäre.

Die Forscher haben zudem Differenzen bezüglich der Einstellungen von Versuchspersonen beider Gruppen nach Beendigung der Strafe beobachtet, wobei diejenigen aus der Gefangenengruppe zu eher negativen Einstellungen neigten. Sie zeigten Frustration über ihre Zuweisung zum Freiheitsentzug, negativere Einstellungen betreffend ihre Tatbegehung, die erhaltene Strafe und die Personen, denen sie dafür die Verantwortung zuschrieben. Während die Autoren der Studie feststellten, dass der Freiheitsentzug zwar keine statistisch signifikant negative Auswirkung auf die Straffälligkeit und Lebensperspektiven der Betroffenen gehabt habe, schien er ihnen zumindest die Einstellung gegenüber Strafen beeinflusst zu haben. So rief die Tatsache, eine (sehr kurze) Freiheitsstrafe verbüßt zu haben

statt gemeinnützige Arbeit abzuleisten, bei den Betroffenen die starke Assoziation einer „unfairen“ Strafe hervor. Die Autoren reduzierten diese Haltung freilich auf einen „Denkfehler“, der auf dem fehlenden Schuldeingeständnis und einer Projektion auf die Behörden beruhe. Aus menschenrechtlicher Perspektive ist dieser Punkt demgegenüber aber ausgesprochen wichtig, weil eine von den Betroffenen als fair wahrgenommene Sanktionspraxis dabei auch ganz unabhängig von den damit erreichbaren Rückfallquoten zentral ist. Es wurde von den Forschern zudem angeführt, dass die positive Entwicklung der Rückfälligkeit in Bezug auf die Inzidenzraten bei den Probanden, die gemeinnützige Arbeit leisten mussten, auf der Tatsache beruhen könnte, dass diese Gruppe eine Wahl gehabt und sich glücklich geschätzt habe, in die Zufallsverteilung einbezogen worden zu sein. Wenn Zufallsverteilungen aber tatsächlich einen solchen Effekt erzielen könnten, dann wäre das bessere Abschneiden von alternativen Sanktionen gegenüber dem Freiheitsentzug in randomisierten Studien – zumindest in gewissem Maße – als ein Artefakt der Forschung zu betrachten. Das ist zumindest das, was die Autoren der systematischen Review nahelegen. Dies würde erneut zeigen, dass randomisierte Forschungsdesign eben nicht, wie gerne behauptet wird, der allen anderen Methoden überlegene „Goldstandard“ sind, sondern dass sie lediglich andere Arten von Fehlern und Problemen aufweisen als es etwa bei quasi-experimentellen Designs der Fall ist. Statt nun aber den Versuch zu unternehmen, eine solche (mögliche) Fehlerhaftigkeit der Ergebnisse zu beseitigen, könnte man auch zu einem anderen Schluss gelangen: Sie kann die Notwendigkeit verdeutlichen, das Fairnessprinzip bei der Anwendung von Sanktionen stärker zu beachten und die Zustimmung der Betroffenen zu ihrer Grundlage zu machen oder diesen bei der Sanktionsauswahl zumindest ein Mitspracherecht, oder sogar ein Wahlrecht, einzuräumen. Dieser Aspekt wird im Folgenden noch detaillierter in Bezug auf den von den Autoren der systematischen Review angesprochenen Hawthorne- bzw. Placeboeffekt erläutert.

V Vergleich von randomisierten zu Matched-Pair-Design-Studien und von Sanktionen zur Nicht-Intervention

Nach dem strikten Paradigma der sog. „evidence-based crime prevention“ und der experimentellen Forschung werden Studien mit einem Matched-Pair-Design im Hinblick auf die Kausalität zwischen den Interventionen und den gemessenen Ergebnissen generell als suboptimal betrachtet. Bei einem Matched Pair Design werden zunächst „Zwillings- oder Mehrlingspaare“ gesucht, die sich in Hinblick auf bestimmte Attribute ähneln, denen eine Einfluss auf die Wahrscheinlichkeit und

Häufigkeit der Begehung von Straftaten zugeschrieben wird. Diese Paare werden dann gezielt getrennt und unterschiedlichen Interventionsgruppen zugewiesen, damit die für relevant gehaltenen Attribute möglichst gleichmäßig auf diese Gruppen verteilt werden, ihr Einfluss sich also bezogen auf die unterschiedlichen Interventionen möglichst nicht in unterschiedlichem Maße bemerkbar macht, so dass deren Wirkung weitgehend unabhängig von diesen Attributen erforschbar werden soll. Bei einem solchen Verfahren liegt es jedoch in der Natur der Sache, dass nur solche Attribute überhaupt Berücksichtigung finden können, von denen man schon auf den Gedanken gekommen ist, dass sie Einfluss auf strafbares Verhalten haben könnten und bislang unbekannte Zusammenhänge nicht berücksichtigt werden können. Dies ist bei einem randomisierten Forschungsdesign anders bei dem auch solche noch nicht bekannten Attribute wenigstens theoretisch annähernd auf die unterschiedlichen Interventionsgruppen verteilt wären, man muss allerdings sagen, dass man deren Einfluss als solchen aber auch mit einer randomisierten Studie nicht erfahren, sondern lediglich statistisch zu neutralisieren versuchen kann. In Anbetracht der geringen Anzahl an randomisierten Studien und natürlichen Experimenten zum Vergleich von Freiheitsstrafen und ambulanten Sanktionen haben die Autoren der Campbell-Review, obwohl sie von der Überlegenheit echter experimenteller Designs überzeugt sind, dennoch 18 quasi-experimentelle Studien mit einem Matched-Pair-Design berücksichtigt. Für die Aufnahme einer Studie in die Review musste diese allerdings mindestens drei oder mehr unabhängige Kontrollvariablen verarbeiten. Aufgrund diverser für die Rückfällig herangezogener Messkriterien in einiger der Studien ergaben sich aus den 23 berücksichtigten Studien (vier randomisierte Experimente, ein natürliches Experiment und 18 Matched-Pair-Design-Studien) insgesamt 27 Vergleiche. In elf von diesen 27 Vergleichen schnitten die ambulanten Sanktionen hinsichtlich der Rückfälligkeit der einbezogenen Personen signifikant besser ab, während diejenigen, die einer freiheitsentziehenden Maßnahme unterzogen worden waren, nur in zwei Studien eine geringere Rückfallquote aufwiesen. In 14 Vergleichen konnten keinerlei signifikante Differenzen festgestellt werden, wobei die ambulanten Sanktionen unterhalb des statistischen Signifikanzniveaus auch in diesen Studien besser abschnitten als die freiheitsentziehenden. Um einem simplifizierenden Ansatz wie einer bloßen „Stimmenauszählung“ von Studienergebnissen, nach dem Muster 11 : 2 für die ambulanten Sanktionen, aus dem Weg zu gehen, bezogen die Autoren nur die fünf Studien in ihre Metaanalyse mit ein, denen sie eine besonders hohe Qualität zuschrieben. In ihren Augen war der Vorteil von ambulanten Sanktionen, der sich aus den Gesamtergebnissen aller 27 Vergleiche ergab, in der Metaanalyse dann kaum noch sichtbar. Obwohl die Unterschiede zwischen den sich aus polizeilichen Akten und Statistiken über Wiederverurteilungen ergebenden Rückfall-

quoten bei ambulanten und freiheitsentziehenden Sanktionen sehr gering waren, wies die Anwendung von nicht-freiheitsentziehenden Sanktionen aber dennoch etwas bessere Ergebnisse auf. Die oben getroffenen Aussagen zu dem schweizerischen Experiment treffen allerdings auch auf die Mehrzahl der anderen Studien zu, welche bezüglich der angewandten Freiheitsstrafen nur sehr kurze Zeiträume betrachteten und ebenso waren die Bezugszeiträume für den Ersatz von Freiheitsstrafen durch ambulante Sanktionen sehr kurz. Dies hängt teilweise auch mit der Tatsache zusammen, dass im Rahmen randomisierter Studien regelmäßig nur einander relativ ähnliche Sanktionsarten miteinander verglichen werden können obwohl es meist interessanter wäre stärker voneinander verschiedene Interventionen und auch eine Non-Intervention miteinander zu vergleichen. Dies ist ein insbesondere bei randomisierten Studien auftretendes, typisches ethisches Problem, da bei einem Vergleich von Sanktionen völlig unterschiedlicher Härte der Grundsatz der Gleichbehandlung verletzt würde, wenn gezielt gleichgelagerte Fälle völlig unterschiedlichen Sanktionen zugeordnet würden, diese Zuordnung gewissermaßen ausgewürfelt würde. Umgekehrt wären aber gerade Vergleiche wie der zwischen Freiheitsstrafen, Bewährung und Nicht-Intervention für die Forschung und eine rationale Kriminalpolitik besonders interessant.⁸⁷ Die Nr. 104 der Europäischen Grundsätze der Bewährungshilfe bestimmt etwa, dass sich die Politik und Praxis der Bewährungshilfe soweit wie möglich auf wissenschaftliche Erkenntnisse stützen sollen und eine Finanzierung gründlicher Forschung gewährleistet sein muss, wobei hier insbesondere auf einen Bedarf an experimentellen Forschungsdesigns hingewiesen wird, deren Ergebnisse als Grundlage für die Politik dienen sollten. Die Nr. 105 der Grundsätze gibt darüber hinaus vor, die bereits bestehenden Normen, Politiken und Praxen auf Basis von fundierten wissenschaftlichen Erkenntnissen und internationalen Standards entsprechender Forschung zu überprüfen und weiterzuentwickeln.

Ein wichtiger Aspekt einer forschungs- und zugleich menschenrechtsorientierten Politik unter den Bedingungen der ersichtlich bestehenden Unsicherheit über die Wirkung unterschiedlicher Sanktionen und kriminalpräventiver Interventionen muss die Anwendung des Prinzips „*primum non nocere*“ sein: vor allem anderen jedenfalls keinen Schaden zuzufügen. Dieses Prinzip entstammt – wie auch die randomisierten Versuche – der Medizin und bezieht sich auf die Notwendigkeit nicht intendierte Wirkungen von medizinischen Interventionen (Risiken und Nebenwirkungen) zu vermeiden als oberstes Prinzip. Es erlangt aber bei Sanktionsentscheidungen eine noch signifikantere Rolle, wenn man bedenkt, dass diese regelmäßig gegen den Willen des betroffenen Individuums angewandt werden,

87 Vgl. Graebisch 2000, S. 271-282.

während die ärztliche Behandlung grundsätzlich auf freiwilliger Basis stattfindet. Das Prinzip „*primum non nocere*“ muss im Strafrecht ähnlich wie das Prinzip „*in dubio pro reo*“ implementiert werden. Jedenfalls solange sich ein intensiver Eingriff in die Rechte der betroffenen Person nicht wenigstens als im Hinblick auf die Prävention von Rückfälligkeit effektiv erweist, darf er nicht erfolgen. Solange zudem verschiedene Sanktionen in experimentellen Studien einen vergleichbaren Effekt erzielen, muss diejenige gewählt werden, deren Eingriffsintensität am niedrigsten ist. Die gegenwärtige Kriminalpolitik ist allerdings offensichtlich weit von einer standardmäßigen Anwendung dieses Grundsatzes entfernt, der juristisch auch aus dem Verhältnismäßigkeitsgrundsatz abgeleitet werden kann. In Anbetracht der Tatsache, dass die Ergebnisse von methodologisch stringenten Studien – wie die oben beschriebenen – eine im Hinblick auf den messbaren Erfolg gegebene Austauschbarkeit verschiedener Sanktionen nahelegen, ist eine Befolgung dieses Grundsatzes und eine entsprechende Absenkung der Eingriffsintensität der Sanktionen unerlässlich. Letzteres trifft insbesondere auf Fälle zu, in denen die angestrebte Auswirkung auf die Prävention von Rückfälligkeit bisher noch nicht einmal nachgewiesen werden konnte. Zieht man die Campbell-Review über ambulante versus freiheitsentziehende Sanktionen als Grundlage für die Kriminalpolitik heran, was erklärtes Ziel dieser Art von Evaluation ist, wird das Ausmaß der gegenwärtig bestehenden Unsicherheit besonders deutlich. So mögen randomisierten Studien eine hohe methodologische Qualität aufweisen, diese bedarf gleichwohl noch weiterer Diskussion und ist keineswegs unumstritten.⁸⁸ Solange die meisten dieser Studien – und es sind bisher nicht viele – aus den Vereinigten Staaten kommen, die ein von den meisten europäischen Ländern völlig abweichendes Kriminaljustizsystem und eine vollkommene andere kriminalpolitische Ausrichtung mit Masseneinhaftierungen haben, können diese Studien nicht als qualitativ hochwertige Grundlage für eine europäische Kriminalpolitik herangezogen werden. Bestenfalls sind sie dazu geeignet, auf Problematiken hinzuweisen, die auch in Europa untersucht werden könnten und sollten. Gleichzeitig geben sie Anlass dazu, die Frage der Kausalität auch in der europäischen Forschung zu berücksichtigen. Sie veranschaulichen zudem – immer wieder – die Tatsache, dass verschiedene Sanktionen mehr oder minder austauschbar sind. Zudem hat insbesondere die experimentelle Forschung immer wieder darauf hingewiesen, dass auch wohlmeinend intendierte Interventionen verheerende Auswirkungen haben können. Letzteres trat bei randomisierten Studien zutage, weil sie im Vergleich zu weniger strikten Kontrollgruppenstudien konfundierende Variablen ausschließen sollen. Ein häufig diskutiertes Beispiel für solche dazwischentretenden Variablen, die auf

88 Zum Ganzen Graebisch 2009.

das Ergebnis einwirken, das daher nicht (alleine) auf die evaluierte Intervention zurückgeführt werden kann, sind Variablen, die mit der Auswahl der Gruppenmitglieder zu tun haben, wenn etwa die stärker motivierten Personen sich für die Experimentalgruppe selbst gemeldet haben oder diejenigen, bei denen die Intervention aussichtsreicher war, von den Programmverantwortlichen für eine zu testende Innovation ausgewählt worden sind. Der Effekt des untersuchten Programmes wird dann im Ergebnis überschätzt. Während dies für streng kontrollierte Studien spricht, wurde jedoch oben bereits angesprochen, dass es auch Gründe gibt, die zur Vorsicht gegenüber randomisierten Designs anhalten. Die Auseinandersetzung mit experimenteller Forschung kann aber dafür sensibilisieren, vermeintlich plausible Kausalzusammenhänge kritisch zu betrachten. Für die Kriminalpolitik bedeutet dies, das Prinzip der ultima ratio und das Verhältnismäßigkeitsprinzip weiterzuentwickeln und die Sanktionspraxis im Hinblick auf diese und auf Grundlage empirischer Forschungsergebnisse zu überprüfen. Die Anwendung von Sanktionen, deren höhere Effektivität gegenüber weniger eingriffsintensiven Sanktionen noch nicht einmal nachgewiesen wurde, ist konsequent abzulehnen. Auch wenn sich dies auf den ersten Blick bescheiden ausnehmen mag, würde es sich bei einer konsequenten Umsetzung als ein riesiger Schritt in Richtung zu einer sowohl auf fundierter Forschung als auch den Menschenrechten beruhenden Kriminalpolitik erweisen.

Berücksichtigt man die oben genannten Bedenken gegenüber naiven Vergleichen zwischen verschiedenen europäischen Rechtssystemen und -kulturen, wird auch deutlich, dass eine einzige randomisierte Studie aus der Schweiz – selbst ungeachtet der aufgeworfenen Fragen über deren Ergebnisse – niemals als eine ausreichende Basis für Rückschlüsse auf die Auswirkungen von Sanktionen in ganz Europa dienen kann. Vielmehr müssen die verschiedenen spezifischen rechtlichen und kulturellen Hintergründe des jeweiligen Landes berücksichtigt werden. Nichtsdestoweniger können die schweizerischen Ergebnisse als ein erster Indikator für die Tatsache betrachtet werden, dass es keine Verbesserung der gegenwärtigen Kriminalpolitik mit sich bringen würde, die Ableistung von gemeinnütziger Arbeit durch kurze Freiheitsstrafen zu ersetzen. Ein solches Fazit hätte allerdings sicherlich auch ohne eine experimentelle Studie gezogen werden können, wenn man das Gebot der Europäischen Grundsätze über die nicht im Gefängnis zu vollziehenden Strafen und Maßnahmen berücksichtigt, Freiheitsentzug durch ambulante Alternativen zu ersetzen. Dies andersherum zu tun, also gemeinnützige Arbeit durch Freiheitsentziehung zu ersetzen, hätte ohnehin offensichtlich einen Rückschritt bedeutet. Solange also gemeinnützige Arbeit von den Betroffenen nicht erwiesenermaßen als beschwerlicher denn kurzzeitiger Freiheitsentzug wahrgenommen wird, muss das Angebot gemeinnütziger Arbeit das Mittel der Wahl bleiben. Beständen

Zweifel im Hinblick auf die Wahrnehmung der Betroffenen – hierzu gibt die vorhandene Forschung jedoch keinen Anlass –, kann ein Wahlrecht der Betroffenen Klarheit schaffen.

Selbst wenn zudem sehr kurze Freiheitstrafen, wie sie in der schweizerischen Studie getestet wurden, zu Effekten führen, die (jenseits der statistischen Signifikanz) mit den Auswirkungen gemeinnütziger Arbeit in anderen Ländern vergleichbar sind, sagt dies selbstverständlich nichts über längere und eingriffsintensivere Haftstrafen aus. In diesem Zusammenhang erscheint es plausibel, anzunehmen, dass langfristige Haftstrafen größere Prisonisierungseffekte bewirken als kurze Freiheitsstrafen. Je länger der Freiheitsentzug andauert, desto wichtiger wird es also aus menschenrechtlicher Perspektive, über mögliche Alternativen nachzudenken. Die Autoren der Campbell-Review weisen diesbezüglich auf die Studie von Smith, Goggin und Gendreau hin, die die Korrelation zwischen der Länge der Inhaftierung und der Rückfälligkeit untersucht haben.⁸⁹ Deren Metaanalyse beinhaltete 117 Studien mit insgesamt 504 Korrelationen. Smith et al. kamen zu dem Schluss, dass die Wahrscheinlichkeit der Rückfälligkeit steigt, je länger die Haftstrafe ist. Sie untersuchten nicht nur die Relation von Rückfallquoten und Haftdauer, sondern auch den Zusammenhang zwischen der Rückfälligkeit und der Sanktionsform, d. h. der Verbüßung einer freiheitsentziehenden Strafe oder einer ambulanten Sanktion. Letzterer Vergleich ergab ungefähr gleiche Ergebnisse für die beiden Sanktionsarten, wobei die Rückfälligkeit nach Verbüßung einer Freiheitsstrafe – je nach der Gewichtung bestimmter Effektstärken – leicht höher war.

Villetta et al. kritisieren die Studie von Smith et al. dafür, vorher existierende Unterschiede zwischen den beiden Gruppen von Verurteilten, der zu freiheitsentziehenden Sanktionen Verurteilten einerseits und der zu ambulanten Sanktionen Verurteilten andererseits, nicht ausreichend berücksichtigt zu haben. Sie bezogen deshalb eine Großzahl der von Smith et al. angeführten Studien nicht mit in ihre Systematic Review mit ein. Diese Diskrepanz in den Herangehensweisen an eine Forschungsübersicht veranschaulicht sehr gut, dass es in der internationalen Forschungsgemeinschaft keineswegs einen einhelligen Konsens über die wichtigsten Kriterien für die Einbeziehung von Studien in eine Metaanalyse oder eine Review gibt. Der Betrachtungsweise der Campbell Collaboration, die experimentellen Forschungsdesigns Exklusivität beimisst, kann konzediert werden, dass randomisierte Designs das Kausalitätsproblem in besonders elaborierter Weise adressieren. Solange experimentelle Studien jedoch auf der Zuweisung von und durch Menschen beruhen, wird das Randomisierungsverfahren immer in gewissem Maße verzerrt bleiben und neben den erwünschten Vorteilen auch Nachteile mit sich bringen.

89 2002.

Mit anderen Worten, ebenso wie Sanktionen neben den intendierten Wirkungen auch nicht intendierte Konsequenzen haben, die die intendierten überlagern und ihre Wirkung aufheben oder in ihr Gegenteil verkehren können, gilt dies auch für experimentelle Forschungsdesigns. Quasi-experimentelle Forschungsdesigns von der Art natürlicher Experimente mögen demgegenüber zwar hinsichtlich des Verständnisses von Variablen, deren Einfluss noch unbekannt ist, weniger verlässlich erscheinen, sie haben aber den Vorteil, dass die Zuteilung zu den Gruppen nicht aktiv durch die Forschung erfolgt, so dass bei dieser auch keine Fehler gemacht werden können. Stattdessen werden bereits bestehende Gruppen untersucht. Der Verzicht auf eine reale und aktive Zuweisung von Sanktionen entsprechend den Kriterien der Forschung bei Zugriff auf ohnehin vorhandene Unterschiede löst zudem die ethischen und rechtlichen Probleme mit Blick auf den Gleichheitsgrundsatz. Während Smith et al. eine große Anzahl an Studien in ihre Review miteinbezogen und damit den Aspekt der Kausalität zu einem gewissen Maße (in problematischer Weise) vernachlässigten, verzichteten Vallettaz et al. in ihrer Metaanalyse auf Bandbreite und Vielfalt von Studien, indem sie lediglich fünf der Einbeziehung für würdig erachteten. Demzufolge sind beide Reviews – aus unterschiedlichen Gründen – für die wichtige Frage des Vergleichs von freiheitsentziehenden und ambulanten Sanktionen von begrenzter Reichweite. Beide Reviews kommen jedoch zu dem eindeutigen Ergebnis, dass ambulante Sanktionen bezüglich der Rückfälligkeit der Betroffenen freiheitsentziehenden Maßnahmen zumindest nicht unterlegen sind. Dieser Umstand mag als Basis für die Beantwortung einer weiteren wichtigen Frage herangezogen werden, nämlich der nach der Sanktion mit der nach Ansicht der Betroffenen geringsten Eingriffsintensität. Die angesichts der unzureichenden Forschungsergebnisse fortbestehende Unsicherheit darf dabei nicht als Vorwand dafür genommen werden, die Tatsache herunterzuspielen, dass ambulante Sanktionen nach dem derzeitigen Erkenntnisstand jedenfalls nicht weniger wirksam sind als freiheitsentziehende Maßnahmen, jedoch mit höchster Wahrscheinlichkeit von den Betroffenen als weniger belastend wahrgenommen werden. Smith et al.⁹⁰ führen hierzu treffend aus, dass Anhänger einer harten Kriminalpolitik an der Qualität der Studien herumkritisieren mögen, die die das Gefängnis diskreditieren, tatsächlich hätten sie jedoch die Anwendung der von ihnen bevorzugten freiheitsentziehenden Sanktionen auf weitaus weniger gesicherte Erkenntnisse, Argumentationen über den gesunden Menschenverstand und unsystematische Forschungsübersichten (narrative reviews) gestützt. Smith et al. weisen zudem darauf hin, dass es in der Literatur zur Verhaltensmodifikation absolut keine Grundlage für die Annahme gibt, dass strafrechtliche Sanktionen mittels Dro-

90 A.a.O. S. 25.

hung, Abschreckung (oftmals die zugrunde liegende These für die Forderung nach Freiheitsstrafen) oder anderweitig die Rückfälligkeit beeinflussen könnten. Dieses Argument deutet wiederum auf die Notwendigkeit von (quasi-)experimentellen Studien hin, die auch die Nichtintervention als Kontrollgruppe in ihre Vergleiche mit aufnehmen. Träfe die sich oftmals aus dem Vergleich verschiedener Sanktionen ergebende Ähnlichkeit von Auswirkungen auch auf den Fall der Nicht-Intervention zu, dann müsste der bestehende Ansatz des Sanktionssystems umso radikaler hinterfragt werden. Dass dies ein Ausgang von entsprechenden Studien sein könnte, ist weniger unwahrscheinlich, als man denken mag. Dies kann etwa an der systematischen Review von Petrosino et al. veranschaulicht werden.⁹¹ Die Review umfasst 29 experimentelle Studien mit 7.304 Jugendlichen und kommt zu dem Schluss, dass das förmliche Verfahren vor dem Jugendstrafericht keine kriminalitätsreduzierende Wirkung zeigte. Vielmehr deuteten fast alle Ergebnisse der Studien auf eine negative Wirkung hinsichtlich der Prävalenz, Inzidenz und Schwere der Kriminalität sowie der von den Betroffenen selbst berichteten Delinquenz hin. Studien, die sich mit dem Vergleich zwischen förmlichem Justizverfahren und Diversionsprogrammen auseinandersetzten, kamen zu dem Ergebnis, dass die Auswirkungen des traditionellen Strafverfahrens im Vergleich zu Diversionsprogrammen wesentlich negativer waren als im Vergleich zu einer Divisionsentscheidung ohne jegliche Intervention. Studien, die ein förmliches Strafgerichtsverfahren mit der Nicht-Intervention verglichen kamen zwar zu Ergebnissen mit einer ähnlichen Tendenz zugunsten der Nicht-Intervention, jedoch war die Diskrepanz zwischen der (positiveren) Wirkung eines Divisionsprogramms gegenüber der Justiz größer als zwischen Nicht-Intervention und Justiz. Die Non-Intervention war also immer noch effektiver als die reguläre Herangehensweise der Jugendstrafjustiz. Dies sollte als Beweis für die Notwendigkeit genügen, die Auswirkungen von Nicht-Intervention gegenüber den Effekten von Sanktionen zu untersuchen, wann immer dies möglich ist, da die Nicht-Intervention oder das Unterbreiten von Angeboten statt zwangsweise auferlegter Maßnahmen aus der menschenrechtlichen Perspektive immer zu präferieren ist und bei vergleichbaren Resultaten den letztgenannten Strategien immer Vorrang gewährt werden muss.

91 Petrosino et al. 2010.

VI Hawthorne und Placebo: Nur im Auge des Betrachters?

Im Jahr 2002 stellte das Ministerkomitee fest, der Einsatz ambulanter Sanktionen und Maßnahmen als Alternative zu freiheitsentziehenden sei gerechtfertigt, da erstere humaner seien und in der Vergangenheit auch weniger gekostet hätten als letztere. Für die Zukunft sah das Ministerkomitee als zusätzlichen Rechtfertigungsgrund die gesteigerte Effektivität ambulanter Sanktionen im Hinblick auf die Rückfälligkeit der Betroffenen.⁹²

Die Autoren der Campbell-Review weisen nachdrücklich darauf hin, dass ambulante Sanktionen deshalb bessere Ergebnisse aufwiesen als freiheitsentziehende Maßnahmen, weil sich die Betroffenen, die ambulanten Sanktionen unterlagen, fairer behandelt fühlten als diejenigen, denen im Rahmen der Studien die Freiheit entzogen worden sei. Tatsächlich zeigte etwa das Schweizer Experiment zur gemeinnützigen Arbeit, dass Betroffene eine solche Wahrnehmung hatten. Vallettaz et al. bezeichnen das bessere Abschneiden der ambulanten Sanktionen daher als einen möglichen Placebo- oder Hawthorne-Effekt, d.h. als einen Effekt, der eher aus der Beobachtung des Ergebnisses einer Intervention als aus dieser selbst resultiert. Die Autoren können dieses Fazit jedoch nicht auf Grundlage der von ihnen so sehr betonten hohen methodologischen Qualität der randomisierten Forschungsdesigns ableiten. Vielmehr nehmen sie einen möglichen Einfluss durch Beobachtereffekte einfach als eine Erklärung an und bedauern zugleich, dass die Durchführung von Doppelblindstudien im Strafrecht im Gegensatz zur Medizin nicht möglich sei. So sei es im Hinblick auf Strafen unmöglich, unterschiedliche Sanktionen zu verhängen, wobei weder die entscheidende noch die betroffene Person wisse, um welche Sanktion es sich jeweils handle. Abgesehen von sehr komplexen und eher geringfügigen Unterschieden z.B. bezüglich einzelner Behandlungsansätze wird die betroffene Person immer wissen, welche Sanktion gerade auf sie angewandt wird, insbesondere wird sie feststellen können, ob gegen sie eine ambulante oder eine freiheitsentziehende Maßnahme angeordnet wurde. Dieses Bedauern der Campbell-Autoren deutet neben vielen anderen Indizien darauf hin, dass die strafrechtliche Sanktionsforschung mit der medizinischen Forschung nicht vergleichbar ist und deshalb ein anderer Ansatz notwendig ist. So mögen randomisierte kontrollierte Doppelblindstudien das Nonplusultra der medizinischen, speziell der pharmakologischen Forschung sein, jedoch müssen sie deshalb noch nicht die beste, geschweige denn die einzig akzeptable Methode für die kriminologische Sanktionsforschung sein. Ein Grund dafür ist die Tatsache, dass

92 Council of Europe 2002.

die Lösung methodologischer Probleme von randomisierten Versuchen, die in der Medizin gefunden wurde, nicht oder zumindest nicht vollständig auf kriminologische Sachverhalte anwendbar ist. In dem Bemühen, sich der medizinischen Forschung anzunähern wird bei experimentellen Studien – insbesondere im Bereich der strafrechtlichen Sanktionierung – oftmals versäumt, die weiterhin bestehenden Unterschiede zwischen der medizinischen und der kriminologischen Forschung zu berücksichtigen. Doppelblindstudien werden, wo sie in der Medizin durchführbar sind, in der Annahme präferiert, dass sie die „wirklichen“, d. h. physiologischen Auswirkungen von Interventionen untersuchen und nicht nur die psychologischen. In Übertragung dieser Herangehensweise erlauben sich die Autoren der Campbell-Review die Schlussfolgerung: „Wenn jedoch innovative »Alternativen« – unabhängig davon, worin diese bestehen und was sie mit sich bringen – , durchgängig bessere Ergebnisse erzielen als traditionelle Sanktionen, dann erscheint es vertretbar, solche Ergebnisse als Hawthorn[e] (oder Placebo)-Effekt zu interpretieren.“⁹³ Hiermit interpretieren sie die Forschungsergebnisse in ihrer Review als Überschätzung des Einflusses ambulanter Sanktionen aufgrund solcher Effekte. Diese Wahrnehmung folgt allerdings einer übermäßigen Anwendung des medizinischen Forschungsmodells im Bereich des Strafrechts. So müssen bei der Anwendung dieses Modells zumindest die Unterschiede zwischen beiden Forschungsfeldern berücksichtigt werden. Die wichtigste Differenz stellt hierbei die Tatsache dar, dass die medizinische Forschung für gewöhnlich Interventionen testet, die später von den Betroffenen auf freiwilliger Basis durchgeführt werden. Das Ziel besteht zwar dabei in der Heilung einer Erkrankung, jedoch entscheiden die betroffenen Personen in einem Akt der Selbstbestimmung, ob sie die entsprechenden Medikamente nehmen bzw. welche sie bevorzugen. Die kriminologische Forschung setzt sich hingegen mit Sanktionen auseinander, die typischerweise gegen den Willen des Betroffenen verhängt werden. Es gibt einige ambulante Sanktionen, oder sogar eine steigende Anzahl solcher, deren Anwendung die Zustimmung der Betroffenen erfordert, jedoch insbesondere im Zusammenhang mit Freiheitsentzug finden die Interventionen regelmäßig im Zwangskontext statt. Stimmt die betroffene Person nicht-freiheitsentziehenden Maßnahmen nicht zu, wäre Freiheitsentzug die zwangsläufige Konsequenz. Dabei ist ein solcher nicht, wie es vereinfachend gerne getan wird, mit einer (ansonsten unheilbaren) Erkrankung zu vergleichen⁹⁴, sondern staatlich veranlasste Übelzufügung, die schlicht unterlassen werden könnte. Sollte sich im Zuge entsprechender Forschung herausstellen – und die Autoren der Campbell-Review weisen auf eine Reihe diese These bestätigende Studien hin –,

93 Killias und Villettaz 2008, S. 33.

94 Vgl. zu diesem ethischen Argumentationsmuster Graebisch 2002, S. 153-199.

dass Sanktionen, die als fair empfunden werden, die Kooperationsbereitschaft der Betroffenen erhöhen, während als unfair wahrgenommene Sanktionen diese Kooperationsbereitschaft zerstören, dann wäre dies keineswegs als ein irrelevantes Artefakt der Forschung zu betrachten. Vielmehr repräsentiert Fairness bei der Messung von Effektivität in der Strafrechtspflege mehr als eine bloße Störvariable. Sie ist nämlich ein sich aus den Menschenrechten ergebendes Prinzip, welches es unabhängig von Fragen effektiver Sanktionierung zu befolgen gilt. Sollte aber nachweisbar sein, dass sich bestimmte Sanktionen sowohl als effektiv als auch (in der Wahrnehmung der Betroffenen) fairer erwiesen, dann ist dies ein wichtiges Argument für eine Anwendung dieser Sanktionen anstelle der als weniger fair empfundenen. Killias und Valletaz gehen weiterhin davon aus, dass „Fairness“ im Hinblick auf randomisierte Studien über strafrechtliche Sanktionen mehr oder weniger als Synonym für „besser als erwartet“ zu betrachten sei. Sollte dies der Fall sein, so könnte es auch die Richtung für die Fortentwicklung des Sanktionensystems aufzeigen. Bisher wurde oftmals angenommen, dass ein hartes Durchgreifen gegen Kriminalität notwendig sei, um auf Seiten des Täters oder der Täterin nicht das Gefühl hervorzurufen, mit der Straftat „davonzukommen“, wenn eine eher milde Sanktion verhängt wird. Den Befunden von Killias und Valletaz zufolge kann jedoch die Verhängung einer Sanktion, die als besser als erwartet empfunden wird, eine Haltung hervorrufen, die der Effektivität der Sanktion sogar zuträglich ist. Dies kann zum Beispiel mit der „Control Balance Theory“ von Charles Tittle erklärt werden. Er geht davon aus, dass das Verhältnis zwischen dem Grad der Kontrolle, dem eine Person ausgesetzt ist und dem Ausmaß, in dem dasselbe Individuum Kontrolle über andere ausüben kann, Einfluss auf die Wahrscheinlichkeit von (bestimmten Formen von) deviantem Verhalten hat.⁹⁵ Das Ergebnis, dass Sanktionen, die den Betroffenen selbst angemessener erscheinen, mit größerer Wahrscheinlichkeit zum Erfolg bezüglich der Rückfälligkeit des Betroffenen führen, geht zudem mit jüngeren Sichtweisen in der Kriminologie auf dem Feld der Resozialisierung einher – insbesondere mit dem „Good Lives Model“.⁹⁶

Obwohl die Unterschiede zwischen ärztlicher Behandlung und strafrechtlichen Sanktionen nicht ignoriert werden dürfen, bleibt anzumerken, dass selbst in der Medizin Placebo-Effekte nicht lediglich als Störfaktor in einer Kette von Kausaleffekten betrachtet werden, sondern vielmehr einen eigenständigen Behandlungsansatz darstellen, der auch bewusst eingesetzt werden kann. Dieser Umstand trifft auf Sanktionen noch umso mehr zu, was besonders deutlich wird, wenn man

95 Für eine empirische Überprüfung dieser Thesen siehe Piquero und Hickman 1999, S. 319-342.

96 Ward und Maruna 2007.

einen Blick auf das ursprüngliche Verständnis des sogenannten Hawthorne-Effekts wirft. Während der Placebo-Effekt der medizinischen Forschung entsprang, wurde der Hawthorne-Effekt erstmals von Henry A. Landsberger in den 1950ern beschrieben, nachdem dieser in den 1920er und 1930er Jahren Experimente in den Hawthorne Elektrizitätswerken durchgeführt hatte, in denen er die Relation zwischen der Produktivität und der Arbeitsumgebung – wie etwa den Einfluss von guten Lichtverhältnissen in den Werkhallen – untersuchte. Es stellte sich heraus, dass die Arbeiter im Zuge der Studie produktiver waren als zuvor, unabhängig davon, ob sie mit dem bestehenden oder einem verbesserten Lichtsystem arbeiteten. Dieser Effekt wurde der Beachtung, die den Arbeitern im Rahmen der Studie zukam, zugeschrieben. Wenn das Ziel der Forschung ist, herauszufinden, ob bessere Lichtverhältnisse die Produktivität der Arbeiter steigert oder nicht, dann kann die den Arbeitern geschenkte Beachtung eine Störvariable hinsichtlich der Ergebnisse darstellen. Geht es aber wie im Bereich der strafrechtlichen Sanktionen – und insbesondere der ambulanten Überwachung – im Rahmen einer Studie darum, das notwendige Minimum an Intervention und Überwachung herauszuarbeiten, dann stellt sich die Situation gänzlich anders dar. Überwachung ist in diesem Fall ein wichtiges Element des Untersuchungsgegenstandes und ebenso ist Fairness ein bedeutsamer Aspekt des Sanktionsrechts und dessen Ethik. Stellt sich heraus, dass Fairness und Anerkennung die Effektivität erhöhen, dann mag das zwar vielleicht nicht als ausreichender Grund angesehen werden, anstelle einer – oder zusätzlich zu einer – besseren Beleuchtung in einer Fabrik den dortigen Arbeitskräften mehr Aufmerksamkeit zukommen zu lassen (obwohl auch das empfehlenswert erscheinen mag), dies stellt sich jedoch im Falle von strafrechtlichen Sanktionen, insbesondere im Hinblick auf Menschenrechtsstandards und das Verhältnismäßigkeitsprinzip, anders dar. Die erlaubte Schwere des Eingriffs ist u. a. nämlich durch die Unumgänglichkeit einer Intervention, die den Betroffenen zwangsweise aufgebürdet wird, begrenzt. Dies hängt mit dem Umstand zusammen, dass der Eingriff in Rechte, besonders in Rechte wie jenes auf persönliche Freiheit, nur als Ultima Ratio, d. h. als letztes Mittel erfolgen darf, wenn mildere Mittel sich als ineffektiv herausgestellt haben.⁹⁷ Folglich müssen Fairness und Zuwendung, welche derzeit noch als Placebo- oder Hawthorne-Effekte betrachtet werden, zukünftig bedeutende Elemente der Sanktionsforschung sein.

Natürlich ist es – wie oben dargelegt – unmöglich, Doppelblindstudien in der vergleichenden Forschung über sich stark voneinander unterscheidende Sanktionen durchzuführen, da die Versuchspersonen schnell realisieren werden, ob sie eine freiheitsentziehende oder eine ambulante Sanktion erhalten haben. Allerdings

97 Siehe Ministerkomitee des Europarates, Empfehlung Nr. R 99 (22), Grundsatz Nr. 1.

ist eine doppelblinde Zuweisung von Sanktionen auch gar nicht notwendig, um Effekte im obigen Sinne zu untersuchen. Es wäre für das selbige Forschungsziel bereits ein Fortschritt, Kontrollgruppen der Non-Intervention einzuführen. Es ist allerdings äußerst schwierig, wenn nicht sogar unmöglich, wirkliche Nichtinterventionsgruppen auf dem Feld der Sanktionenforschung einzubeziehen, da es zumindest immer strafrechtliche Ermittlungen geben wird, die auf die Feststellung der Tat und eines oder einer Tatverdächtigen zielen und damit unter Umständen schon als Intervention wahrgenommen werden können (ähnlich wie die ärztliche Untersuchung vor einem Behandlungsversuch). Würde jedoch einen Schritt später im Verfahren die Strafverfolgung durch eine Diversionsentscheidung (in Deutschland der Staatsanwaltschaft) gänzlich gestoppt und die Einstellung des Verfahrens nicht mit irgendwelchen Bedingungen, Auflagen oder Weisungen verbunden, würde dies einer Nonintervention bzw. einer placeboähnlichen Intervention bereits sehr nahekommen.

Es lässt sich darüber streiten, ob verschiedene Sanktionsarten im Rahmen eines experimentellen Forschungsdesigns in Form einer bewussten Zuweisung ähnlicher Fälle zu unterschiedlichen Maßnahmen wie etwa einer Freiheitsstrafe einerseits und einer ambulanten Sanktion andererseits getestet werden dürfen⁹⁸ und sollten.⁹⁹ Sollte es aber rechtlich und ethisch vertretbar sein, den Grundsatz der Gleichheit vor dem Gesetz um der Forschungsergebnisse willen zu verletzen bzw. um durch diese spätere Rechtssetzung und Rechtspraxis zu beeinflussen, und sollte es ferner als legitim betrachtet werden, Personen nach dem Zufallsprinzip einer Freiheitsstrafe oder einer ambulanten Sanktion zuzuweisen, dann sollte es auch möglich sein, eine Kontrollgruppe der Nicht-Intervention einzuführen. Dagegen könnte allenfalls angeführt werden, dass Freiheitsstrafen und Nicht-Intervention zu unterschiedlich seien, um berechtigterweise als Gruppen einer einzigen Studie zu fungieren, auch wegen des Schutzes potenzieller Opfer im Falle schwerer Straftaten. Für diesen Fall könnte die Durchführung von zwei Studien angedacht werden: eine, die die Freiheitsstrafe mit einer ambulanten Sanktion vergleicht und – wenn die ambulante Sanktion in dieser ersten Studie zumindest nicht schlechtere Resultate aufwies als die Freiheitsstrafe – eine weitere, die einen Vergleich zwischen der ambulanten Sanktion und Nicht-Intervention zieht. Bei einer derart konzipierten Studie mag ein Nachteil sein, dass die zugeordneten Individuen erkennen können, welcher Art von Behandlung sie ausgesetzt werden bzw. – im Falle der Nicht-Interventionsgruppe – welcher Behandlung sie entgehen konnten. Aber die

98 Vgl. zu dieser Problematik Graebisch 2000, S. 271-282; dies. 2007, S. 193-204; dies. 2009.

99 Dazu zuletzt Graebisch 2014b.

mit einer solchen Argumentation implizit zugrunde gelegte Annahme einer Notwendigkeit, Interventionen doppelblind zuzuweisen, dürfte aus dem Verständnis resultieren, dass Beobachtungseffekte nur dann untersucht werden könnten, wenn die betroffene Person, die einer Placebo-Behandlung oder einer Nicht-Intervention zugewiesen wurde, hiervon nichts weiß, sondern von einer „echten“ Sanktionierung ausgeht. Dies dürfte mit der Assoziation einer täuschend echt aussehenden Pille, die jedoch keinen Wirkstoff enthält, verbunden sein. Eine solche Wahrnehmung würde jedoch bedeuten, die Eigenart solcher Effekte zu verkennen und gering zu schätzen. Wie es schon lange aus der medizinischen Forschung bekannt ist, können Placebo-Medikamente nämlich durchaus einen Effekt haben, selbst wenn die betroffene Person weiß, dass sie lediglich ein Placebo erhalten hat.¹⁰⁰ Oder um das Martinsons zugeschriebene Zitat¹⁰¹ mittels einer veränderten Betonung umzuschreiben: von nichts wirkt in Bezug auf Kriminalprävention („nothing works in crime prevention“) hin dazu, die Möglichkeit anzuerkennen, dass nichts in Bezug auf Kriminalprävention eine Wirkung haben könnte („nothing works in crime prevention“).

Der Vergleich von ambulanten Sanktionen und Nicht-Interventionen erscheint gegenüber dem bloßen Vergleich der ambulanten mit freiheitsentziehenden Sanktionen dann als noch wichtiger, wenn es um das Problem des Net-Widening geht, welches im Folgenden behandelt werden soll. Umso mehr sich nämlich ambulante Sanktionen dem Verdacht ausgesetzt sehen müssen, als Ersatz für eine Nicht-Intervention und nicht als Ersatz für freiheitsentziehende Maßnahmen angewandt zu werden, desto mehr muss sich die Forschung auch mit dem Vergleich zu Nicht-Interventionen beschäftigen.

VII Natürliches Experiment: Ausnutzung zufallsgesteuerter Zuweisung bei Gerichten

Um ethischen Vorbehalten gegenüber einer randomisierten Zuweisung von Sanktionen (einschließlich der Nicht-Intervention) aus dem Weg zu gehen, können natürliche Experimente als eine weitere Möglichkeit gelten, mit der dem Placeboproblem mittels vergleichender Forschung auf den Grund gegangen werden kann. Die bekannten Unterschiede in der Praxis strafrechtlicher Sanktionsentscheidungen bei vergleichbaren Fällen, die z. B. auf die verschiedenen Persönlichkeiten der ent-

100 Park und Kovi 1965, S. 336-345.

101 Dazu Graebisch 2011, S. 137-147.

scheidenden Richter und Richterinnen oder unterschiedliche regionale Gerichtskulturen zurückgehen, werden regelmäßig als gewichtiges Argument für die Ethik der Zufallszuordnung bei Strafsanktionen zu Forschungszwecken herangezogen. Dabei wird angeführt, dass Unterschiede gleicher Intensität wie sie bereits in der bestehenden Sanktionierungspraxis zu finden seien und die aus Gründen, die mit einer Zuweisung zu Forschungszwecken nichts zu tun haben, aber offenbar hingenommen werden, den Maßstab dafür darstellen könnten, inwieweit dann auch eine Zufallszuweisung unter der Geltung des Gleichheitsgrundsatzes vertretbar sein könne, zumal diese als Ausgangspunkt dafür dienen könne, diese Unterschiede perspektivisch entsprechend den so gewonnenen Erkenntnissen der Forschung folgend zu reduzieren. Unterschiede in einem dem bereits existierenden Spektrum von Ungleichbehandlung vergleichbarer Fälle entsprechenden Ausmaß sollten dieser Argumentation zufolge mittels einer Zufallszuordnung verteilt werden dürfen, um Vergleichsgruppen konstruieren und ermitteln zu können, welche der verschiedenen Interventionen angesichts der (exakten) Forschungsergebnisse die vorzugswürdigste sei.¹⁰² Diese Argumentation könnte im besten Falle dann als akzeptabel betrachtet werden, wenn die bestehenden Unterschiede nach der Durchführung des Experiments verringert und somit kohärente Entscheidungen getroffen würden, die zudem auf den Ergebnissen der Studie basierten. Es gibt jedoch keine Anhaltspunkte dafür, dass dergleichen nach einer experimentellen Studie geschehen würde.¹⁰³ Gleichzeitig gibt es keine ethischen Vorbehalte gegenüber der Verwendung von divergierenden Sanktionsentscheidungen verschiedener Richter und Richterinnen als Ausgangspunkt für eine Analyse, die im methodologischen Rahmen eines natürlichen Experiments erfolgt. Letzteres drängt sich insbesondere in Fällen auf, bei denen Beschuldigte per Zufall bestimmten Spruchkörpern zugeordnet und abweichende Straftendenzen erwartet werden können. Es ist überraschend, dass dieser auf der Hand liegende Forschungsansatz bisher so selten verfolgt wurde.

Ein Beispiel für eine solche Studie ist die von Green und Winik.¹⁰⁴ Die Autoren analysierten die Fälle von 1.003 Angeklagten, denen Rauschgiftdelikte vorgeworfen und die per Zufallszuordnung über eine Zeitspanne von vier Jahren neun Richtern oder Richterinnen zugewiesen worden waren. Die Entscheidungen der Richter und Richterinnen unterschieden sich erheblich im Hinblick auf die Dauer der Freiheitsstrafe und die Bewährungszeit. Die am wenigsten strenge Gruppe von Richtern verurteilte 23 Prozent der Angeklagten zu Freiheitsstrafen, wohingegen die

102 Federal Judicial Center 1981.

103 Dazu bereits Graebisch 2004, S. 266-283.

104 Green und Winik 2010, S. 357-387.

strengste Richtergruppe in 65 Prozent der Fälle auf eine Haftstrafe entschied. Die Autoren fanden jedoch keinen systematischen Zusammenhang zwischen Eigenschaften der Angeklagten, die mit Kriminalität in Verbindung gebracht werden und dem Spruchkörper, dem sie zugeordnet worden waren, so dass die Gruppen von Verurteilten als annähernd gleich zusammengesetzt gelten können. Es war daher in diese Weise möglich, im Hinblick auf etwa 40 Prozent der Verurteilten Schlussfolgerungen zu der Effektivität von Freiheitsstrafen im Vergleich zu Bewährungsstrafen in dem Maße anzustellen, in dem diese Verurteilten unterschiedlich behandelt worden waren. Es wurde mit anderen Worten möglich, den Effekt der zufälligen Verdoppelung oder fast Verdreifachung von durch ein Gericht angeordneten Gefängnisstrafen für diese Gruppe von Straffälligen zu untersuchen. Im Ergebnis fand die Studie von Green und Winik, dass sich weder Freiheitsentzug noch Bewährung als statistisch signifikante Abschreckungsmaßnahmen für eine Rückfälligkeit während der auf die Studie folgenden vier Jahre – gemessen an neuerlichen Festnahmen – erwiesen. Die Autoren folgerten: „Unter ansonsten gleichbleibenden Bedingungen ist zu erwarten, dass der im Mittelmaß der untersuchten Gruppe angesiedelte Angeklagte, der Inhaftierung und Bewährung erfahren hat, mit in etwa der selben Wahrscheinlichkeit rückfällig werden wird wie ein Angeklagter, der ohne Bestrafung oder Überwachung entlassen würde.“¹⁰⁵ Das Ergebnis dieser Studie zeigt wiederum eindringlich, dass neben Sanktionsalternativen auch Alternativen zu Sanktionen in Betracht zu ziehen sind.

105 Green und Winik 2010, S. 375.