

Entwicklung und Stand der empirischen Strafzumessungsforschung

Bernd-Dieter Meier

1. Fragestellung

Einer der wissenschaftlichen Schwerpunkte des Lehrstuhls Strafrecht, Strafprozessrecht und Kriminologie ist die empirische Strafzumessungsforschung. Gegenstand dieses Forschungszweigs ist die mit erfahrungswissenschaftlichen Methoden betriebene Untersuchung der Frage, wie die normativen, insbesondere in § 46 StGB geregelten Vorgaben für die Strafzumessung in der Praxis umgesetzt werden. Etwas konkreter formuliert geht es um die Aufklärung, von welchen Gesichtspunkten sich die Gerichte bei der Zumessung von Strafe tatsächlich leiten lassen und inwieweit sich hierin die wenigen, in der Regel durch weite Beurteilungsspielräume und Ermessen gekennzeichneten Vorgaben des Gesetzgebers widerspiegeln. Die Strafzumessungsforschung ist Teil der allgemeiner gefassten Sanktionsforschung, die sich mit der empirischen Analyse der staatlichen Reaktionen auf abweichendes Verhalten befasst, also mit den Strafen, den Maßregeln der Besserung und Sicherung, den Nebenfolgen und den Alternativen zur Strafe, und die nicht nur die gerichtliche Anordnungspraxis, sondern auch die Vollstreckung und den Vollzug sowie die intendierten und die nicht-intendierten Wirkungen der Sanktionen in den Blick nimmt.¹ Mit dieser Ausrichtung knüpfen die am Lehrstuhl durchgeführten Untersuchungen an eine Forschungstradition an, die in Deutschland eine vergleichsweise lange Geschichte hat: Schon am Göttinger Lehrstuhl von *Heinz Schöch* wurde empirische Strafzumessungsforschung betrieben;² erinnert sei aber auch an

1 *Meier*, Sanktionsforschung, in: *Schneider* (Hrsg.), Internationales Handbuch der Kriminologie, Bd. 1: Grundlagen der Kriminologie, Berlin 2007, S. 971-1010.

2 Grundlegend *Schöch*, Strafzumessungspraxis und Verkehrsdelinquenz: Kriminologische Aspekte der Strafzumessung am Beispiel einer empirischen Untersuchung zur Trunkenheit im Verkehr, Stuttgart 1973; hierauf aufbauend *Meier*, Die Bestrafung von Rückfalltätern in der Bundesrepublik Deutschland und in den USA: Eine rechtstatsächliche und rechtsvergleichende Untersuchung zu § 48 StGB, Göttingen 1983; *Hoppenworth*, Strafzumessung beim Raub: Eine empirische Untersuchung der Rechts-

die „Studien über die Strafzumessungspraxis der deutschen Gerichte“, die schon 1931 von dem in nationalsozialistisches Unrecht verstrickten Strafrechtler und Kriminologen *Franz Exner* vorgelegt worden waren.

Wenn Gerichte in einer Strafsache über das Strafmaß zu entscheiden haben, stehen sie vor zwei Herausforderungen: Sie müssen entscheiden, welche Umstände im Einzelfall für die Zumessung der Strafe bestimmend sein sollen, und sie müssen entscheiden, in welches konkrete Strafmaß diese Umstände umgesetzt werden sollen. Nach § 46 Abs. 1 StGB und der heute herrschenden Meinung kommt es für beide Entscheidungen wesentlich auf die Schuld des Täters an; präventive Überlegungen dürfen nur in dem durch die Schuldschwere gezogenen Spielraum in Ansatz gebracht werden. Für die erste Entscheidung bedeutet dies, dass strafzumessungsrelevant vor allem solche Umstände sind, die sich zum tatbestandlichen Unrecht und seiner Vorwerfbarkeit in Beziehung setzen lassen. In der Regel handelt es sich dabei um Umstände, die das Ausmaß des Erfolgsunwerts kennzeichnen, beispielsweise um die Schwere der Verletzungen oder den Wert des entwendeten Guts, sowie um Umstände mit Bezug zum Handlungsunwert, also um die Beweggründe und die Ziele, die der Tat zugrunde gelegen haben, aus dem Vorleben des Täters um die Vorstrafen, aus seinem Nachtatverhalten das Geständnis. Für die zweite Entscheidung bedeutet dies, dass der qualitativen Aussage über die Schwere der festgestellten Schuld (beispielsweise mittelschwere Verletzungen, vorangegangene Provokation, keine Vorstrafen, Geständnis schon in der ersten Vernehmung) ein quantitativer Wert zugeordnet werden muss, der die Schuldschwere im Vergleich zu anderen, weniger gewichtigen oder schwereren Taten verdeutlicht und dessen Grenzen sich nach unten aus dem schon und nach oben aus dem noch schuldangemessenen Beurteilungsspielraum ergeben, der vom Strafraumen zur Verfügung gestellt wird. Eine Orientierung liefern die Eckpunkte des gesetzlichen Strafraumens: Das gesetzliche Mindestmaß bezeichnet diejenige Strafschwere, die für die denkbar leichtesten Fälle verhängt werden muss, das gesetzliche Höchstmaß entsprechend die Grenze für die denkbar schwersten Fälle.

In der Öffentlichkeit stößt diese Weite des trichterlichen Ermessens auf Befremden. Befürchtet wird, dass in diesem normativ nur wenig regu-

folgenzumessung bei Verurteilungen wegen Raubes nach allgemeinem Strafrecht und nach Jugendstrafrecht, München 1991; *Verrel*, Schuldfähigkeitsbegutachtung und Strafzumessung bei Tötungsdelikten: Eine empirische Untersuchung zur Bedeutung des psychowissenschaftlichen Sachverständigen im Strafverfahren, München 1995.

lierten Bereich sachfremde Erwägungen angestellt und Entscheidungen getroffen werden, die die Schwere der Schuld nicht zutreffend widerspiegeln. Sachfremd wäre es etwa, wenn in die Entscheidung Umstände einfließen würden, die sich zum tatbestandlichen Unrecht und seiner Vorwerfbarkeit nicht in Beziehung setzen lassen, Umstände wie die Sympathie einer Richterin für die politische Gesinnung des Angeklagten oder die Abneigung eines Richters gegen Personen mit Migrationshintergrund. Befürchtet wird in jüngerer Zeit umgekehrt auch, dass in die Entscheidung nicht oder nicht ausreichend die „richtigen“, aus der gesellschaftlichen Perspektive für relevant gehaltenen Umstände einfließen. Die Befürchtung betrifft insbesondere die unzureichende Berücksichtigung menschenverachtender Beweggründe oder Ziele, die der Tat zugrunde gelegen haben; mit drei Novellierungen von 2015, 2021 und 2023 sind vom Gesetzgeber etliche Motive hervorgehoben worden, die bei der Strafzumessung berücksichtigt werden sollen („rassistische, fremdenfeindliche, antisemitische, geschlechtsspezifische, gegen die sexuelle Orientierung gerichtete oder sonstige menschenverachtende“). Quer zu alledem wird schon seit langer Zeit immer wieder die Befürchtung geäußert, dass sich die richterliche Einordnung in den Strafraum weniger am Gesetz als an der lokalen Justizkultur orientiere, was zu regionalen Ungleichheiten führe. Schon in der Studie von *Franz Exner* findet sich im Jahr 1931 der Satz: „Für den Dieb, der vor einen Hamburger Richter kommt, ist ... die Aussicht, mit einer höheren Strafe belegt zu werden, viermal größer als für seinen Genossen im Stuttgarter Bezirk.“³ Auch wenn sich aus verfassungsrechtlicher Sicht gegen regionale Ungleichheiten bei der Strafzumessung keine Einwände erheben lassen,⁴ ist dies ein Thema, das seither immer wieder erörtert wird und im Jahr 2018 sogar zu einer intensiven Diskussion in der strafrechtlichen Abteilung des 72. Deutschen Juristentags geführt hat.⁵

Die am Lehrstuhl betriebene Strafzumessungsforschung will zu diesen Fragen einen empirischen, erfahrungswissenschaftlichen Beitrag leisten. Im Mittelpunkt steht die Bestandsaufnahme, welche Entscheidungen von den

3 *Exner*, Studien über die Strafzumessungspraxis der deutschen Gerichte, Leipzig 1931, S. 49.

4 BVerfGE 1, 332 (345 f.); vertiefend *Meier*, Regionale Justizkulturen in der Strafrechtspraxis: ein Problem für den Rechtsstaat?, in: Dessecker/Egg (Hrsg.), Justizvollzug und Strafrechtsreform im Bundesstaat, Wiesbaden 2011, S. 31-49.

5 *Kaspar*, Sentencing Guidelines versus freies richterliches Ermessen – Brauchen wir ein neues Strafzumessungsrecht?, Gutachten C zum 72. Deutschen Juristentag, München 2018.

Gerichten in welchen Sachverhaltskonstellationen getroffen werden und welche Umstände dabei tatsächlich bedeutsam gewesen sind. Mit der Frage nach den „tatsächlich bedeutsamen“ Strafzumessungsfaktoren wird ein Blick auf die richterliche Entscheidungspraxis geworfen, der von den in den schriftlichen Urteilen genannten Begründungen unabhängig ist, indem er nach den statistischen Zusammenhängen zwischen einzelnen denkbaren Strafzumessungstatsachen und der Sanktionsschwere fragt. Wenn die verschiedenen Ausprägungen eines potentiellen Strafzumessungsfaktors im Strafmaß zu signifikanten Unterschieden führen, ist die Vermutung nahelegend, dass es sich hierbei um einen Umstand handelt, den die Gerichte typischerweise berücksichtigen. Mit multivariaten Analysen, in die mehrere potentielle Strafzumessungsfaktoren eingehen, lässt sich dabei auch die Gleichzeitigkeit des gleich- und gegenläufigen Wirkens unterschiedlicher Faktoren statistisch abbilden (beispielsweise die bereits erwähnten mittelschweren Verletzungen in Kombination mit vorangegangener Provokation, keinen Vorstrafen und einem Geständnis schon in der ersten Vernehmung). Durch den Vergleich der statistischen Befunde mit den in den Urteilsgründen genannten Umständen lässt sich ein auf die untersuchte Gruppe (nicht auf Einzelfälle) bezogener Eindruck von der Wahrheit und Vollständigkeit der Urteilsbegründungen gewinnen. Hiervon ausgehend kann in einem zweiten Schritt danach gefragt werden, inwieweit sich in diesen Entscheidungen die rechtlichen Maßgaben widerspiegeln, insbesondere ob es Hinweise auf sachfremde Einflüsse oder Erledigungsroutinen gibt, die von diesen Maßgaben nicht gedeckt sind, womit sich ein Eindruck von der Richtigkeit der Strafzumessungsentscheidungen gewinnen lässt.⁶ Dieser Versuch, richterliche Entscheidungen mithilfe eines alternativen, erfahrungswissenschaftlichen Ansatzes nachzuvollziehen und der Idee nach zu kontrollieren, löst auf der Seite der Kontrollierten typischerweise große Skepsis und Ablehnung aus. Um solchen Einwänden begegnen zu können, ist es notwendig, sich genauer mit dem methodischen Vorgehen bei diesen Untersuchungen zu beschäftigen und zu fragen, inwieweit sich gerichtliche Entscheidungen und ihr Zustandekommen in empirischen Untersuchungen abbilden lassen.

6 Vgl. Meier, Rückfalltäter (Fn. 2), S. 176 ff.

2. Methodik

a) Vorgehen bei der Datenerhebung

Die Warnung eines Praktikers „Bewerten Sie nicht ein Urteil, dessen Entstehung Sie nicht mitbekommen haben“, weist darauf hin, dass die erste und wichtigste Frage, die im Zusammenhang mit empirischer Strafzumessungsforschung zu klären ist, dem Vorgehen bei der Datenerhebung gilt. Die Entstehung eines Urteils lässt sich methodisch am besten mit einer Beobachtung verfolgen, wobei insbesondere an die Beobachtung von Hauptverhandlungen zu denken ist. Für die Eignung der Methode spricht der die Durchführung der Hauptverhandlung gestaltende Mündlichkeitsgrundsatz, wonach dem Urteil nur der mündlich vorgetragene und erörterte Prozessstoff zugrunde gelegt werden darf; die Mündlichkeit soll gerade für Dritte die Transparenz der Entscheidungsfindung gewährleisten. Für die Methode spricht zudem, dass sich mit ihr auch solche Informationen gewinnen lassen, über die nicht gesprochen wird und die auch nicht verschriftlicht werden, die aber dennoch auf die Strafmaßfindung Einfluss nehmen können, etwa nonverbale Signale wie die stockenden Äußerungen des Angeklagten in der Vernehmung, das Weinen der Zeugin oder die erwähnte augenscheinliche Sympathie der Vorsitzenden für den Angeklagten. Beobachtungen haben indes einen unüberwindbaren Nachteil: Derjenige Verfahrensabschnitt, in dem das Urteil tatsächlich „entsteht“, nämlich die Beratung, ist der Beobachtung durch Dritte entzogen; Beratungen sind nicht öffentlich (§ 193 Abs. 1 und 2 GVG), so dass sich über den Gang der Beratung, die ausgetauschten Argumente und die Frage, warum sich welches Argument durchgesetzt hat, mit der Methode der Beobachtung kein Aufschluss gewinnen lässt. Auch die Befragung der Beteiligten scheidet als Datenerhebungsmethode aus, da über den Gang der Beratung und die Abstimmung nicht gesprochen werden darf (§ 43, 45 Abs. 1 Satz 2 DRiG).

In der Strafzumessungsforschung wird vor diesem Hintergrund durchgängig mit Aktenanalysen gearbeitet, also mit der systematischen Auswertung der in den Verfahrensakten gesammelten Informationen über den Gang des Verfahrens und die getroffenen Entscheidungen. Auf die Akten-einsicht besteht für Forschende bei Vorliegen der gesetzlichen Voraussetzungen ein Anspruch (§ 476 Abs. 1 und 2 StPO).

Aktenanalysen haben für die Strafzumessungsforschung zahlreiche Vorteile. Der wesentliche Vorteil gegenüber der Beobachtung und der Befragung liegt darin, dass Akten eine zuverlässige Zusammenstellung sämtlicher

verfahrens- und entscheidungsrelevanten Informationen über den jeweiligen Fall beinhalten. Die Zuverlässigkeit und Vollständigkeit des von Polizei und Justiz zusammengetragenen Materials folgt aus der Kommunikationsfunktion der Akten: Anhand der Akten können sich die Verfahrensbeteiligten, aber beispielsweise auch Amtsnachfolger oder Urlaubsvertreter, über den jeweils aktuellen Kenntnisstand des Gerichts oder der Behörde und den Stand des Verfahrens informieren.⁷ Dies gilt auch für die Ergebnisse von Vernehmungen oder den Gang der Hauptverhandlung, die in Protokollen festgehalten und zur Akte genommen werden. Die in den Akten zusammengetragenen Informationen eignen sich zwar nicht, um die Lebenswirklichkeit der Verfahrensbeteiligten zu erschließen (z.B. zu der Frage, in welchen wirtschaftlichen Verhältnissen der Angeklagte lebt, insbesondere wie hoch sein monatliches Nettoeinkommen ist); sie bilden eine Wirklichkeit „eigener Art“, die durch die Perspektive der Strafverfolgungsorgane geprägt ist.⁸ Für die Strafzumessungsforschung liegt hierin jedoch kein Nachteil, da es für sie gerade darum geht festzustellen, wie die Justiz mit den ihr vorliegenden Informationen umgeht, welche Informationen sie überhaupt für entscheidungsrelevant hält, nach welchen Prinzipien sie diese Informationen aufklärt und mit welchem Ergebnis die Informationen im Strafzumessungsvorgang verarbeitet werden. Ein nicht uninteressanter Befund der empirisch-kriminologischen Forschung kann so beispielsweise gerade darin bestehen, dass die für die Strafzumessung relevante Höhe des Nettoeinkommens (§ 40 Abs. 2 Satz 2 StGB) in der Praxis häufig nicht aufgeklärt, sondern geschätzt wird (§ 40 Abs. 3 StGB), was beim Fehlen von Anknüpfungstatsachen für die Schätzung einer vom Bundesverfassungsgericht für unzulässig erklärten Schätzung „ins Blaue hinein“ gleichkommt.⁹

b) Notwendige Standardisierungen

„Jeder Fall ist anders.“ Der in der Justiz oft zu hörende Satz macht in seiner augenscheinlichen Trivialität auf ein Problem aufmerksam, das als das zentrale Problem der empirischen Strafzumessungsforschung bezeich-

7 Vgl. Dölling, Probleme der Aktenanalyse in der Kriminologie, in: Kury (Hrsg.), Methodologische Probleme in der kriminologischen Forschung, Köln u.a. 1984, S. 265 (270 f.); Leuschner/Hüneke, Möglichkeiten und Grenzen der Aktenanalyse als zentrale Methode der empirisch-kriminologischen Forschung, MschrKrim 2016, 464 (466 ff.).

8 Dölling, Aktenanalyse (Fn. 7), S. 270.

9 Vgl. Kolsch, Sozioökonomische Ungleichheit im Strafverfahren, Berlin/Münster 2020, S. 301 ff., 426 ff.; BVerfG NStZ-RR 2016, 46.

net werden kann: Wie können Fälle, die völlig unterschiedlich sind, angemessen kodiert, also in statistisch prüfbare Einheiten übersetzt und so miteinander vergleichbar gemacht werden? Dass die Fälle miteinander verglichen werden müssen, wenn man das Strafmaß verstehen will, liegt auf der Hand und lässt sich am Beispiel des Zitats von *Exner* aus dem Jahr 1931 verdeutlichen: Wenn Diebe in Hamburg härter bestraft werden als Diebe in Stuttgart, kann der einfache Grund hierfür ja darin liegen, dass die abgeurteilten Diebstähle unterschiedlich sind. Dies wiederum kann zwei Ursachen haben: Die begangenen Diebstähle können unterschiedlich schwer wiegen, beispielsweise weil der Wert der entwendeten Sachen in Hamburg höher ist als in Stuttgart, oder die Staatsanwaltschaften können von ihren Einstellungsmöglichkeiten in unterschiedlicher Weise Gebrauch gemacht haben, beispielsweise können in Hamburg mehr Bagatelldiebstähle nach § 153 StPO eingestellt worden sein, was zur Folge hätte, dass von den Gerichten in Hamburg nur die schwereren Diebstähle abgeurteilt werden. Die Beantwortbarkeit beider Fragen setzt voraus, dass die begangenen Diebstähle angemessen durch Variablen abgebildet werden können, damit den Fragen empirisch weiter nachgegangen werden kann. Was macht also die Schwere eines Diebstahls aus? Wie lässt sich die Schwere von Delikten messen?

aa) Deliktschwere

Es liegt auf der Hand, dass sich die Vielgestaltigkeit der von der Justiz verhandelten Lebenssachverhalte in statistischen Untersuchungen nicht abbilden lässt. Es ist unumgänglich, die aus der Akte ersichtlichen Fälle auf einige wenige Merkmalsdimensionen zu reduzieren, um sie für die empirische Analyse handhabbar zu machen. Diese unvermeidbare Reduzierung der Komplexität der Lebenssachverhalte ist im Grunde nicht problematisch. Sie entspricht dem Vorgehen, das auch die Gerichte anwenden, wenn sie die festgestellten Strafzumessungstatsachen (bei einem Körperverletzungsdelikt z.B. die mittelschweren Verletzungen) quantifizieren, indem sie sie in ein konkretes Strafmaß übersetzen. Allerdings ist in der empirischen Forschung bei der Übersetzung des angeklagten Sachverhalts in statistisch prüfbare Variablen Vorsicht geboten: Die gebildeten Variablen müssen den Gütekriterien der empirischen Forschung entsprechen, insbesondere müssen sie valide und reliabel sein. Gemeint ist damit, dass die gebildeten Variablen den erfassten Lebenssachverhalt in genau denjenigen Dimensio-

nen erfassen müssen, die für die empirisch nachzuvollziehende Strafzumessungsentscheidung des Gerichts wesentlich sind, und dass sie von den die Zuordnung im Einzelfall durchführenden Personen unabhängig sein müssen, also bei Wiederholung der Messung durch andere Personen zu denselben Ergebnissen führen müssen. Eindeutig zu beantworten muss also beispielsweise die Frage sein, ob der ausgeschlagene Zahn und der gebrochene Mittelfinger vergleichbare „mittelschwere Verletzungen“ sind oder ob ihnen ein unterschiedliches Gewicht zukommt.

In empirischen Untersuchungen zur Strafzumessung spielt die Deliktschwere eine zentrale Rolle. Aus rechtlicher Sicht scheint das auf den ersten Blick zwar angreifbar zu sein, denn rechtlich kommt es nicht auf die Schwere der begangenen Taten, sondern auf die Schuld des Täters, also auf das Ausmaß der Vorwerfbarkeit an (§ 46 Abs. 1 Satz 1 StGB). Bezugspunkt der Vorwerfbarkeit ist jedoch das tatbestandliche Unrecht, wobei strafzumessungsrechtlich das Maß des zurechenbar verwirklichten Erfolgsunwerts die entscheidende Rolle spielt.¹⁰ In empirischen Untersuchungen eine Variable zu bilden, in der sich das Erfolgsunrecht widerspiegelt, um diese dann statistisch mit dem Strafmaß in Verbindung zu setzen, ist deshalb im Grundsatz richtig.

Die Schwierigkeiten beginnen jenseits dieser allgemeinen Aussagen. Zum einen müssen Kategorien gefunden werden, mit denen sich die Deliktschwere im Einzelfall zutreffend abbilden lässt. Hierfür müssen Skalen entwickelt werden, die dem jeweiligen Untersuchungsgegenstand angemessen sind. Für den Schweregrad des Taterfolgs bei der Körperverletzung lässt sich beispielsweise in Anlehnung an die Schmerzensgeldtabellen eine 6-stufige Skala bilden, die von Bagatellverletzungen bis zu schweren Verletzungen reicht.¹¹ Für die Erfassung der Schwere von Missbrauchsabbildungen hat sich international die 10-stufige COPINE-Skala durchgesetzt, die von nichtsexualisierten, aber unangemessenen Bildern bis zur Darstellung von Gewalthandlungen gegen ein Kind reichen.¹² So sinnvoll und hilfreich

10 Schäfer/Sander/van Gemmeren, Praxis der Strafzumessung, 6. Aufl., München 2017, Rn. 1178 ff.

11 Kolsch, Ungleichheit (Fn. 9), S. 194; vgl. aber auch Weiterer, Die Strafzumessung bei Vollrausch (§ 323a StGB): Eine rechtsdogmatische und empirische Untersuchung, jur. Diss. Hannover 2022, Berlin 2024, S. 149 ff.

12 Taylor/Quayle, Abusive Images of Children and the Internet: Research from the COPINE Project, in: Cooper et al. (Hrsg.), Medical, Legal and Social Science Aspects of Child Sexual Exploitation: A Comprehensive Review of Pornography, Prostitution and Internet Crimes, St. Louis 2005, S. 257 ff.

derartige Instrumente sind, so deutlich ist allerdings auch, dass sie in empirischen Untersuchungen für die Erfassung der Deliktschwere nur dann ausreichen, wenn nur ein einziger Deliktsbereich untersucht werden soll, also beispielsweise nur Körperverletzungen oder nur Missbrauchsabbildungen. In der justiziellen Wirklichkeit geht es dagegen um ganz unterschiedliche Kriminalitätsphänomene, die von den Gerichten in Strafmaßentscheidungen umgesetzt werden müssen. In der Justizwirklichkeit ist die Frage nicht so sehr, ob der ausgeschlagene Zahn oder der gebrochene Mittelfinger schwerer wiegt, sondern die Frage ist, ob der ausgeschlagene Zahn oder der Ladendiebstahl schwerer wiegt, bei dem ein handelsübliches Smartphone entwendet worden ist. Für Erfassung derartiger Entscheidungen müssen in empirischen Untersuchungen abstraktere Skalen gebildet werden, bei denen in die Bewertung beispielsweise die in den Strafraumen zum Ausdruck kommenden Vorwertungen des Gesetzgebers, die Schwereinschätzungen von Richterinnen und Richtern und / oder die Ergebnisse aus Bevölkerungsumfragen eingehen.¹³ Dass derartige Skalen wiederum den Anforderungen der Validität und Reliabilität genügen müssen, ist offensichtlich.

Aber selbst wenn eine Skala gebildet wird, die die Deliktschwere angemessen abbildet, sind die methodologischen Probleme damit noch nicht abschließend gelöst. Zum anderen stellt sich nämlich das Problem, dass in der justiziellen Wirklichkeit häufig nicht nur ein Delikt angeklagt und abgeurteilt wird, sondern mehrere Delikte, die zueinander in Tateinheit oder Tatmehrheit stehen, was sich bei der Strafzumessung ganz unterschiedlich auswirkt (§§ 52, 54 StGB). Auch diese Situation der mehrfachen Tatbestandsverwirklichung muss in einer empirischen Untersuchung angemessen abgebildet werden. In der Forschungspraxis wird dabei meist in der Weise verfahren, dass den Regeln für die Gesamtstrafenbildung folgend nur das schwerste Delikt bzw. die Einsatzstrafe kodiert wird (vgl. § 54 Abs. 2 Satz 2 StGB) und die übrigen, tatmehrheitlich konkurrierenden Delikte bzw. Einzelstrafen nur nach ihrer Anzahl erfasst werden. Bei der statistischen Auswertung führt diese Vereinfachung zu erheblichen Erleichterun-

13 Vgl. etwa *Schöch*, Ist Kriminalität normal? Probleme und Ergebnisse der Dunkelfeldforschung, in: Göppinger/Kaiser (Hrsg.), *Kriminologie und Strafverfahren: Neuere Ergebnisse zur Dunkelfeldforschung in Deutschland*, Bericht über die XVIII. Tagung der Gesellschaft für die gesamte Kriminologie vom 9. bis 12. Oktober 1975 in Freiburg, Stuttgart 1976, S. 211 (214, 226); ausführlich *Villmow*, Schwereinschätzung von Delikten: Schicht- und altersspezifische Einstellungen sowie Einstellungen von Tätern und Opfern bei 14- bis 25jährigen männlichen Probanden einer südbadischen Kleinstadt, Berlin 1977.

gen, aber es ist offensichtlich, dass sich die Komplexität der richterlichen Gesamtstrafenbildung hierin nur unzureichend abbildet.

bb) Strafhärte

Die Probleme der richtigen, also der validen und reliablen Erfassung der strafgerichtlichen Entscheidung in kodierten Variablen stellen sich nicht nur auf der Seite der potentiellen Einflussfaktoren, sondern auch auf der Seite der abhängigen Variablen, dem Ergebnis des Verfahrens. Wenn untersucht werden soll, wie sich die verschiedenen Ausprägungen eines Strafzumessungsfaktors im Strafmaß auswirken, ist es vergleichsweise einfach festzustellen, ob sich der Umstand auf die Anzahl der verhängten Tagessätze oder der ausgeurteilten Monate Freiheitsstrafe auswirkt. Auch ob die verhängte Freiheitsstrafe unter bestimmten Umständen eher zur Bewährung ausgesetzt wird (§ 56 StGB) oder ob anstelle der Geldstrafe eine Freiheitsstrafe verhängt wird (§ 47 StGB), erschließt sich relativ schnell. Mit diesen Feststellungen ist jedoch noch nichts über die Strafhärte gesagt, so wie sie sich aus der Sicht des Gerichts darstellt oder wie sie vom Verurteilten empfunden wird. Wenn man nur die drei Stufen betrachtet Geldstrafe / Freiheitsstrafe mit Bewährung / Freiheitsstrafe ohne Bewährung, ist es keineswegs sicher, dass eine hohe Geldstrafe weniger schwer wiegt als eine geringe Freiheitsstrafe, deren Vollstreckung zur Bewährung ausgesetzt wird. Auch dass eine hohe, aber ausgesetzte Freiheitsstrafe leichter wiegt als eine geringe, aber vollstreckte Freiheitsstrafe, ist nicht ausgemacht; man denke nur an die Konsequenzen bei einem Bewährungswiderruf. Dass sich die von den Gerichten intendierte und von den Verurteilten empfundene Strafhärte einfachen Gleichungen entzieht, zeigt sich auch an dem Mechanismus der Gesamtstrafenbildung: Zwei 10-monatige Freiheitsstrafen werden in Deutschland nicht zu einer 20-monatigen Freiheitsstrafe addiert, sondern sie werden zu einer Gesamtfreiheitsstrafe zusammengezogen, die sich zwischen 10 Monaten und einer Woche auf der einen und 19 Monaten auf der anderen Seite bewegt (§§ 39, 54 Abs. 1 Satz 2, Abs. 2 Satz 1 StGB). Die Gesamtstrafenbildung beruht auf der Vorstellung, dass die Strafwirkung – das Strafleid des Verurteilten – mit der Strafdauer überproportional zunimmt: Das Strafleid wachse nicht linear, sondern progressiv an; dem

müsse mit dem Verbot der Addition der Einzelstrafen (§ 54 Abs. 2 Satz 1 StGB) entgegengewirkt werden.¹⁴

Unabhängig davon, ob die These von der progressiven Zunahme des Strafleids der Wirklichkeit entspricht oder ob es sich um eine juristische Setzung handelt, besteht Anlass zu der Annahme, dass die drei Strafkategorien Geldstrafe / Bewährung / Freiheitsstrafe im deutschen Recht keine kontinuierlich verlaufende Strafschwereskala bilden, die für die empirische Ermittlung von Unterschieden und Veränderungen im Strafmaß genutzt werden kann. Aus rechtlicher Sicht gilt dies schon deshalb nicht, will sich die Auswahl zwischen den drei Strafkategorien anders als es § 46 StGB vermuten lässt, nicht nur an der Schwere der Schuld, sondern auch an den Bedürfnissen der Prävention orientiert (§§ 47, 56 StGB). Um die Unterschiede und Veränderungen in einer empirischen Untersuchung in einer einheitlichen Skala zusammenfassen zu können, muss deshalb ein Instrument entwickelt werden, in dem unter Vernachlässigung der Präventionsinteressen allein die Härte der Rechtsfolge zum Ausdruck kommt, die das Gericht zum Ausgleich der Schuld verhängt (§ 46 Abs. 1 Satz 1 StGB). Im Kriminologischen Forschungsinstitut Niedersachsen (KFN), das bis 2013 mit der Leibniz Universität durch einen Kooperationsvertrag verbunden war,¹⁵ wurde in den 1990er Jahren ein solches Instrument entwickelt, das es ermöglichte, die Strafhärte einer von den Gerichten verhängten Sanktion unabhängig von der Strafart auf einer Skala zwischen 0 und 100 einzuordnen.¹⁶ Das Instrument war hochplausibel konzipiert, aber es hatte einen Nachteil: Die Skala war nur für Geld- und Freiheitsstrafen bis zu zwei Jahren nach allgemeinem Strafrecht anwendbar. Hohe Freiheitsstrafen und nach Jugendstrafrecht verhängte Sanktionen konnten nicht einbezogen werden. Diese Defizite nahm die Lehrstuhlmitarbeiterin *Mira Behnken* im Jahr 2020 zum Anlass, nach dem Modell des KFN ein weiteres Instrument zu entwickeln, das einerseits hohe Freiheitsstrafen abdeckt und es anderer-

14 Schäfer/Sander/van Gemmeren, Strafzumessung (Fn. 10), Rn. 1199; krit. hierzu Fröhlich, Das Asperationsprinzip: Eine systematische Analyse, jur. Diss. Hannover, 1996.

15 Wissenschaftliche Kommission Niedersachsen, Strukturanalyse der außeruniversitären Forschung in Niedersachsen, Hannover 2019, S. 189.

16 Oswald, Psychologie des richterlichen Strafens, Stuttgart 1994, S. 103 ff.; Langer, Staatsanwälte und Richter: Justitielles Entscheidungsverhalten zwischen Sachzwang und lokaler Justizkultur, Stuttgart 1994, S. 187 ff.

seits erlaubt, die jugend- und die erwachsenstrafrechtlichen Sanktionen zu verbinden und auf einer einheitlichen Strafhärteskala abzubilden.¹⁷

c) Mixed Methods

Der Satz „Jeder Fall ist anders“ macht noch auf ein zweites Problem aufmerksam. Wenn es mithilfe von Delikts- und Sanktionshäreskalen möglich ist, bei unterschiedlichen Fällen unterschiedliche Strafmaßentscheidungen miteinander zu vergleichen, dann geht die hierin liegende Reduzierung von Komplexität unvermeidbar mit Informationsverlusten einher, die die Aussagekraft der empirischen Ergebnisse und damit auch den wissenschaftlichen Wert von Untersuchungen schmälern. Für die unzureichende Erfassung der Gesamtstrafenbildung wurde hierauf bereits hingewiesen, aber es gilt auch für die eigentlich interessierende Frage nach den Einflussfaktoren auf die Strafzumessung. Nicht ausgeschlossen ist es bei der statistischen Gegenüberstellung von Strafzumessungsfaktoren und Strafhärte nämlich, dass die Entscheidung im Einzelfall auch von Umständen beeinflusst wird, die bei der statistischen Gegenüberstellung nicht berücksichtigt werden, beispielsweise wenn bei einem Diebstahl kein egoistisches, sondern ein atypisches altruistisches Tatmotiv vorliegt. Solche Informationen lassen sich zwar ebenfalls in Variablen erfassen und statistisch ausdrücken, aber die hohe Bedeutung, die ihnen im Einzelfall zukommen kann, kann sich in der Menge der statistisch analysierten Zusammenhänge verlieren und damit zu einer Schwächung des Aussagehalts einer Untersuchung beitragen.

Diesem Problem kann in der empirischen Forschung mit einem „mixed methods“ Ansatz abgeholfen werden, bei dem die standardisierende, quantifizierende Herangehensweise mit einer qualitativen Herangehensweise kombiniert wird, bei der die Einzelfallanalyse im Mittelpunkt steht. Für die qualitative Auswertung von Verfahrensakten gibt es – anders als für die quantitative Auswertung – keine etablierte, in der Forschung allseits akzeptierte Methodik, obwohl die Notwendigkeit ergänzender qualitativer Analysen seit langem thematisiert wird;¹⁸ in der am Lehrstuhl betriebenen

17 Behnsen, Strafhärteskalierung. Ein empirischer Lösungsansatz im Bereich hoher Freiheits- und Jugendstrafen, MschrKrim 2020, 45 ff.

18 Dölling, Aktenanalyse (Fn. 7), S. 278; Schumann, Qualitative Forschungsmethoden in der Kriminologie, in: Kury (Hrsg.), Entwicklungstendenzen kriminologischer Forschung: Interdisziplinäre Wissenschaft zwischen Politik und Praxis, Köln u.a. 1986, S. 147 ff.; Leuschner/Hüneke (Fn. 7), MschrKrim 2016, 476 ff.

Forschungspraxis reicht die Spannweite etwa von der bloßen Veranschaulichung statistischer Befunde durch die Schilderung von Einzelfällen¹⁹ bis zur vergleichenden Auswertung einzelner Strafzumessungsbegründungen, die die Schwierigkeiten der Tatgerichte mit der Umsetzung höchstrichterlichen Maßgaben deutlich macht.²⁰ Um den Wildwuchs zu ordnen, haben *Helena Schüttler* und *Carolin Neubert* vom KFN jüngst versucht, diejenigen Analyseschritte zu systematisieren, die bei qualitativen Aktenanalysen gegangen werden sollten. Sie schlagen vor, die einzelnen Aktenbestandteile im ersten Schritt nach Form, Kontext, Inhalt und Bedeutung zu analysieren und im zweiten Schritt nach der Entwicklung des Untersuchungsgegenstands und seinen Veränderungen innerhalb der Akte sowie nach der Bedeutung der Akte im jeweiligen organisationalen Rahmen zu fragen.²¹ Bezogen auf die Strafzumessungsforschung lässt sich mit diesem Vorgehen beispielsweise herausarbeiten, von welchen Akteuren in welchem Verfahrensstadium welche für die Strafzumessung relevanten Umstände in das Verfahren eingebracht worden sind; anzunehmen ist, dass die Konstruktionsleistungen der Verfahrensbeteiligten bei den vergleichsweise einfach festzustellenden, aus juristischer Sicht „objektiven“ Umständen wie der Deliktschwere oder der Vorstrafenbelastung andere sein werden als bei den „subjektiven“ Umständen wie dem Motiv oder der entlastenden Bedeutung des Geständnisses. Solche qualitativen Analysen können die Strafzumessungsforschung erheblich bereichern; die Entwicklung einer spezifischen Methodik steht allerdings noch am Anfang.

3. Ergebnisse

Welche Ergebnisse haben die am Lehrstuhl durchgeführten Untersuchungen erbracht? Was war neu? Was war rechtspolitisch interessant? Strafzumessungsfragen wurden und werden am Lehrstuhl in ganz unterschiedlichen Zusammenhängen empirisch ausgeleuchtet, etwa im Zusammenhang

19 Etwa *Meier*, Rückfalltäter (Fn. 2), S. 197 ff.

20 *Weiterer*, Strafzumessung (Fn. 11), S. 165 ff., 274 ff.

21 *Schüttler/Neubert*, Ordnungsversuche in einem unübersichtlichen Feld – Was leisten qualitative Aktenanalysen in der kriminologischen Forschung?, in: *Jukschat/Leimbach/Neubert* (Hrsg.), *Qualitative Kriminologie, quo vadis? Stand, Herausforderungen und Perspektiven qualitativer Forschung in der Kriminologie*, Weinheim 2021, S. 88 (102 ff.).

mit der Frage, wie sich das (hohe) Alter²² oder der sozioökonomische Status des Angeklagten²³ auf die Strafzumessung auswirken. Eine Studie zur Strafzumessung bei verbaler Gewalt sowie eine gemeinsam mit dem KFN durchgeführte Untersuchung zu den Auswirkungen der gesetzlichen Änderungen in § 46 Abs. 2 Satz 2 StGB sind in Vorbereitung. Drei abgeschlossene Projekte sollen hier genauer vorgestellt werden.

Johanna Kunze untersuchte die Frage, welche Umstände die Strafzumessung bei Mordverurteilungen beeinflussen. Grundsätzlich führt die Verurteilung wegen Mordes zwar ohne jedes Ermessen zu lebenslanger Freiheitsstrafe (§ 211 Abs. 1 StGB), aber die spätere Aussetzung des Strafrests kann bei besonderer Schwere der Schuld zeitlich hinausgeschoben werden (§ 57a Abs. 1 Satz 1 Nr. 2 StGB), und bei Vorliegen gesetzlicher Strafmilderungsgründe wie verminderter Schuldfähigkeit (§§ 21 StGB) kann die absolute Strafe nach unten geöffnet werden (§ 49 Abs. 1 Nr. 1 StGB). *Kunze* wertete in ihrer Untersuchung sämtliche Verurteilungen aus, die in Deutschland in den Jahren 2013 und 2014 nach allgemeinem Strafrecht wegen vollendeten Mordes ergangen waren. Insgesamt lagen ihrer Auswertung 185 Urteile zugrunde, in denen 196 Angeklagte wegen der Tötung von 221 Personen verurteilt worden waren.²⁴

Aus der Vielzahl der Befunde soll hier nur auf die Ergebnisse zur besonderen Schwere der Schuld eingegangen werden. Für 42 Angeklagte (21,4 %) wurde die besondere Schwere der Schuld im Urteil festgestellt. Es zeigte sich, dass dieser Erschwerungsgrund vor allem dann angenommen wird, wenn mehrere Personen getötet werden, wenn es weitere Personen gibt, bei denen die Tötung versucht worden ist, wenn zwei Mordmerkmale verwirklicht worden sind (niedrige Beweggründe und Heimtücke), wenn es sich um Zufallsoffer handelt, wenn der Angeklagte rücksichtslos vorgegangen ist und wenn er eine erhebliche Zahl an Vorstrafen (zehn oder mehr) aufweist.²⁵ Die ersten beiden Faktoren machen dabei deutlich, dass die Gerichte mit der Feststellung der besonderen Schwere der Schuld dem

22 *Poltrock*, Gleichbehandlung oder altersentsprechende Differenzierung: Brauchen wir ein besonderes „Altersstrafrecht“, Mönchengladbach 2013, S. 374 ff.

23 *Kolsch*, Ungleichheit (Fn. 9), S. 394 ff.

24 *Kunze*, Strafzumessung bei Mord: Eine empirische Untersuchung zur Rechtsfolgenentscheidung bei Verurteilungen wegen vollendeten Mordes nach allgemeinem Strafrecht in den Jahren 2013 und 2014, Berlin u.a. 2021, S. 161.

25 *Kunze*, Strafzumessung (Fn. 24), S. 231, 245; vgl. aus jüngerer Zeit hierzu auch *Goldmann/Zähringer*, Die Umstände von Gewicht bei der Feststellung der besonderen Schwere der Schuld: Eine Analyse von Mordurteilen im Zeitraum 2000-2020, Mschr-Krim 2023, 100 ff.

Umstand Rechnung tragen, dass bei der lebenslangen Freiheitsstrafe keine Gesamtstrafenbildung möglich ist; eine härtere Strafe als die lebenslange Freiheitsstrafe gibt es in Deutschland nicht (§ 54 Abs. 1 Satz 1 StGB). Ebenfalls sachgerecht erscheint es, wenn die Gerichte mit der Feststellung die besondere Verwerflichkeit der Tat, die in der Verwirklichung von zwei Mordmerkmalen zum Ausdruck kommt, das rücksichtslose Vorgehen und die hohe Zahl an Vorstrafen berücksichtigen. Auch dies sind gängige Strafschärfungsgründe. Auffällig ist allerdings der Befund, dass bei der Entscheidung auch die Täter-Opfer-Beziehung verarbeitet und die besondere Schuldschwere vor allem dann festgestellt wird, wenn sich die Tat gegen ein Zufallsopfer, also vor allem gegen unbekannte Personen gerichtet hat. Die aus der Istanbul-Konvention von 2011 bekannte Forderung, dass es gerade umgekehrt strafschärfend berücksichtigt werden sollte, wenn sich die Tat gegen derzeitige oder frühere Partnerinnen gerichtet hat (Art. 46) – was heute als geschlechtsspezifisches Motiv in § 46 Abs. 2 StGB sogar gesetzlich normiert ist –, spiegelt sich in diesem Befund nicht wider.

Einer besonderen Fallkonstellation ging *Mira Behnsen* in ihrer Arbeit zur Strafzumessung bei Neonatizid nach. Sie knüpfte mit ihrer Untersuchung an das Datenmaterial an, das das KFN in einem Projekt zu Tötungsdelikten erhoben hatte, die an Kindern unter 6 Jahren begangen und in den Jahren 1997 bis 2006 bekannt geworden waren.²⁶ In 199 Fällen, etwa einem Drittel der Fälle, handelte es sich um die Tötung von Neugeborenen. In 100 dieser Neonatizide kam es zu einer Verurteilung der Mutter; in drei Fällen wurden neben der Mutter weitere Personen (Kindsvater, Großeltern) verurteilt. Zwei Drittel der Täterinnen wurden nach allgemeinem Strafrecht, ein Drittel nach Jugendstrafrecht verurteilt. *Behnsen* interessierte sich ausschließlich für die Strafmaße, die die Gerichte bei den Müttern ausgeurteilt hatten. Rechtspolitisch war diese Frage deshalb von Interesse, weil der Gesetzgeber 1998 den früheren § 217 StGB aufgehoben hatte, der Mütter, die ihr nichteheliches Kind in oder gleich nach der Geburt getötet hatten, privilegierte, indem er für die Taten Freiheitsstrafe von 3 bis 15 Jahren, in minder schweren Fällen von 6 Monaten bis 5 Jahre vorsah.²⁷

Auch hier seien nur einige wenige Ergebnisse herausgegriffen. *Behnsen* stellte fest, dass für die Strafzumessung den gesetzlichen Strafraumen die

26 Höynck/Behnsen/Zähringer, Tötungsdelikte an Kindern unter 6 Jahren in Deutschland: Eine kriminologische Untersuchung anhand von Strafverfahrensakten (1997-2006), Wiesbaden 2015.

27 Behnsen, Strafzumessung bei Neonatizid, Wiesbaden 2023, S. 82 f., 85.

größte Bedeutung zukommt, in den die Gerichte die Tat nach der Aufhebung von § 217 StGB a.F. einordnen, ob sie die Tat also als Totschlag (§ 212 Abs. 1), als minder schweren Fall des Totschlags (§ 213 StGB) oder als Mord (§ 211 Abs. 1 StGB) bewerten. Lässt man die tatbestandliche Einordnung unberücksichtigt, erwiesen sich für die Strafzumessung vor allem drei Umstände als maßgeblich: die Tötung mehrerer Kinder (Mehrfachneonazid), die besondere Verantwortlichkeit, die die Gerichte angesichts des Ausfalls anderer Schutzpersonen der Mutter beimaßen, und die verminderte Schuldfähigkeit (§ 21 StGB), die sie der Mutter für die Ausnahmesituation der Entbindung zubilligten. Soweit nicht auf die Strafhärte, sondern auf die Zahl der ausgeurteilten Monate als abhängige Variable abgestellt wurde, erwies sich zusätzlich als bedeutsam, ob die Täterinnen Schwierigkeiten in der Erwerbstätigkeit hatten: Täterinnen mit Problemen in diesem Bereich wurden etwas härter bestraft also solche ohne Auffälligkeiten, was sich als Hinweis auf ein vom Gericht als problematisch empfundenes Sozialverhalten interpretieren ließ.²⁸ Die Ergebnisse waren insgesamt unspektakulär, machten aber deutlich, dass sich die entscheidenden Weichenstellungen nach der Aufhebung von § 217 StGB von der Rechtsfolgenreise auf die Tatbestandsseite und dabei insbesondere auf die Annahme eines minder schweren Falls (§ 213 StGB) verlagert hatten.

Mit einem ganz anderen Deliktsbereich beschäftigte sich *Anna Weiterer*. Sie untersuchte, an welchen Gesichtspunkten sich die Strafzumessung bei Vollrausch (§ 323a StGB) orientiert. Wegen Vollrauschs wird bestraft, wer sich durch alkoholische Getränke oder andere Mittel in einen Rausch versetzt und in diesem Zustand eine rechtswidrige Tat begeht, für die er wegen Schuldunfähigkeit nicht bestraft werden kann. Die im Rausch begangene Tat wird vom BGH als objektive Strafbarkeitsbedingung interpretiert; § 323a StGB gilt als abstraktes Gefährdungsdelikt. Die Rauschtat darf deshalb bei der Strafzumessung nur indirekt berücksichtigt werden: Bestraft wird der Rausch; die tatbezogenen Merkmale der im Rausch begangene Tat (Art, Umfang, Schwere und Gefährlichkeit, Folgen) dürfen nur als Indizien für die Gefährlichkeit des Rauschs gewertet werden.²⁹ Für die Untersuchung der Frage, wie die Tatgerichte mit diesen Maßgaben umgehen, wertete *Weiterer* 419 Verfahrensakten aus. Dabei handelte es sich

28 Behnsen, Strafzumessung (Fn. 27), S. 191, 213 ff.

29 BGHSt 1, 124 (125); 16, 124; 23, 375 (376); 38, 356 (361); Hecker, in: Eser et al., Schönke/Schröder StGB – Kommentar, 30. Aufl., München 2019, § 323a Rn. 29.

um eine Zufallsauswahl aus den insgesamt 1.011 Verurteilungen, die für 2018 im Bundeszentralregister ausgewiesen werden.³⁰

Weiterer stellte fest, dass bei Verurteilungen nach § 323a StGB der größte Einfluss auf die Strafzumessung von der Art und Schwere der Rauschat ausgeht: Mit zunehmender Deliktschwere steigt nicht nur die Wahrscheinlichkeit, dass anstelle einer Geldstrafe eine Freiheitsstrafe verhängt wird, sondern desto höher fallen auch das Strafmaß bei der Freiheitsstrafe und die Strafhärte aus. Allein mit der Schwere der Rauschat wird mehr als ein Drittel der Varianz des Strafmaßes erklärt. Weitere Umstände wie die Höhe des verursachten Schadens, Wahnvorstellungen des Täters und die Anzahl der Vorstrafen wegen Vollrauschs wirken sich zwar ebenfalls strafschärfend aus, haben aber ein geringeres Gewicht. Andererseits wirkten sich die Umstände des Sich-Berauschns, insbesondere die Unterscheidung zwischen vorsätzlichem und fahrlässigem Sich-Berauschn, kaum auf das Strafmaß aus. In den schriftlichen Urteilsbegründungen spiegelten sich die statistischen Befunde dabei nicht wider. Die Gerichte werteten in den Urteilsgründen vor allem ein Geständnis bzw. Reue und Schuldeinsicht strafmildernd, was statistisch aber bedeutungslos war. Die qualitative Auswertung der schriftlichen Urteilsgründe machte darüber hinaus deutlich, dass die vom BGH für maßgeblich gehaltene Unterscheidung von tat- und täterbezogenen Merkmalen, deren Berücksichtigung bei der Strafzumessung entweder zulässig oder unzulässig ist, in der tatgerichtlichen Praxis fast immer offen gelassen und zum Teil sogar missachtet wird.³¹ Die Schlussfolgerung ist deshalb naheliegend, dass die „Indizkonstruktion“ des BGH für die tatrichterliche Praxis keine entscheidungsleitende Bedeutung hat; eine Reform des Tatbestands und insbesondere eine deutliche Verengung des gegenwärtigen Strafrahmens von § 323a StGB erscheinen rechtspolitisch angezeigt.

4. Einordnung

Der Überblick lässt erkennen, dass es bei der empirischen Strafzumessungsforschung nicht nur um akademische Spielereien geht, sondern dass den Ergebnissen auch eine rechtspolitische Relevanz zukommen kann. Zu

30 Weiterer, Strafzumessung (Fn. 11), S. 145 f., 172 f.; Kurzfassung in: Bliesener et al. (Hrsg.), Kriminalität und Kriminologie im Zeitalter der Digitalisierung, München-Gladbach 2023, S. 441 ff.

31 Weiterer, Strafzumessung (Fn. 11), S. 232 ff., 254 ff., 266 ff., 298 ff.

hohe Erwartungen dürfen dabei nicht angelegt werden; die Ergebnisse der am Lehrstuhl durchgeführten Untersuchungen haben – was einer größeren Verbreitung immer abträglich ist – keinen hohen Aufmerksamkeitswert. Verallgemeinernd lassen sich dennoch zwei Ergebnisse festhalten: Zum einen hat sich gezeigt – und lässt sich nun empirisch belegen –, dass in den schriftlichen Urteilsgründen nicht zwingend diejenigen Umstände genannt werden, die das Strafmaß tatsächlich bestimmt haben. Es gibt Konvergenzen und Divergenzen, die nicht allein darauf beruhen, dass die Gerichte von Gesetzes wegen schriftlich nur diejenigen Umstände zu benennen brauchen, die für die Zumessung der Strafe „bestimmend“ gewesen sind (§ 267 Abs. 3 Satz 1 StPO). Berücksichtigt werden muss auch, welche kommunikative Funktion die schriftlichen Urteilsgründe haben: Sie wollen nicht die Öffentlichkeit – und schon gar nicht etwaige Forschende – darüber in Kenntnis setzen, welche Überlegungen tatsächlich das Strafmaß getragen haben, sondern sie wollen allein die Verfahrensbeteiligten darüber informieren, welche Umstände aus der Vielzahl theoretisch denkbarer Einflussfaktoren dem normativen Programm des Strafzumessungsrechts entsprechend ausgewählt und berücksichtigt worden sind. Adressat ist dabei auch nicht der Angeklagte: Er und die Saalöffentlichkeit werden in der mündlichen Urteilsbegründung in einer lebensnäheren Sprache, vielleicht auch mit Auslassungen und Rücksichtnahmen (vgl. § 268 Abs. 2 Satz 3 StPO, § 54 Abs. 1, Satz 2, Abs. 2 JGG), über die wesentlichen Erwägungen des Gerichts in Kenntnis gesetzt. Adressaten des schriftlichen Urteils sind in erster Linie die Staatsanwaltschaft und die Verteidigung sowie, wenn Rechtsmittel eingelegt werden (vgl. § 267 Abs. 4 StPO), das Rechtsmittelgericht; ihre Aufgabe ist es, die Übereinstimmung der Entscheidung mit dem normativen Programm zu prüfen. Gerade wegen der Fehleranfälligkeit und potentiellen Aufhebbarkeit der Entscheidung werden die Urteilsgründe von den Gerichten im Zweifel eher knapp gehalten und auf das Wesentliche beschränkt. Dass in der empirischen Forschung Divergenzen mit den tatsächlich wirkenden Strafzumessungsfaktoren ermittelt werden, ist damit zwar erwartbar, erzeugt bei Fachleuten aber keine Aufregung.

Zum anderen hat sich in den empirischen Forschungen gezeigt, dass sich ein systematischer Einfluss sachfremder Gesichtspunkte, die sich zum tatbestandlichen Unrecht und seiner Vorwerfbarkeit nicht in Beziehung setzen lassen, in Strafzumessungsentscheidungen kaum nachweisen lässt. In den meisten Untersuchungen haben sich bei der Strafmaßentscheidung nur solche Merkmale statistisch durchgesetzt, die sich mit dem normati-

ven Programm unter Schuld- oder Präventions Gesichtspunkten erklären lassen, während sich Gesichtspunkte, die in der Öffentlichkeit gelegentlich im Zusammenhang mit Einzelfällen diskutiert werden, bei der statistischen Betrachtung bislang nicht als ausschlaggebend erwiesen haben. In der öffentlichen Debatte können zwei Aspekte für Aufregung sorgen: die Geschlechtsvariable und der Migrationshintergrund, etwa wenn geäußert wird, Straftaten gegen Frauen oder gegen Personen mit Migrationshintergrund würden zu milde, Straftaten, die von Personen mit Migrationshintergrund begangen werden, dagegen zu hart geahndet. Für beide Annahmen hat sich in den Untersuchungen kein Beleg finden lassen. Auch der sozio-ökonomische Status des Angeklagten wirkt sich im Übrigen nicht erst bei der Strafzumessung einflussnehmend aus, sondern schon bei der vorgelagerten Entscheidung der Staatsanwaltschaft über die Erledigung des Verfahrens: Bei wohlhabenden Beschuldigten werden die Verfahren häufiger nach § 153a Abs.1 StPO gegen Zahlung einer Geldauflage eingestellt, während finanziell schlechter gestellte Beschuldigte *ceteris paribus* eher angeklagt werden.³² Lediglich in der Untersuchung von *Kunze*, in der zusätzlich auch nach der Relevanz von personenbezogenen Daten gefragt wurde, die aus dem Handbuch der Justiz ermittelbar waren, ließ sich beobachten, dass die Strafmaßentscheidungen mit dem Alter des Vorsitzenden und der Dauer seiner Diensttätigkeit variierten: Mit zunehmendem Alter wurden die Urteile milder.³³

Andere methodische Ansätze und andere Operationalisierungen mögen zu anderen Erkenntnissen führen. Die Deliktschwere wird in aktuelleren Untersuchungen zum Teil anders operationalisiert³⁴, und auch die Sanktionshärte wird teilweise anders erfasst.³⁵ Die am Lehrstuhl durchgeführten

32 *Kolsch*, Ungleichheit (Fn. 9), S. 275 ff., 351 ff.

33 *Kunze*, Strafzumessung (Fn. 24), S. 355 f.

34 Vgl. *Bliesener et al.*, Prozess- und Wirkungsevaluation der NRW-Initiative „Kurve kriegen“ – Abschlussbericht, Kiel 2015, S. 50 ff., 228 ff.; dieser Index wurde allerdings schon in der Arbeit von *Weiterer*, Strafzumessung (Fn. 11), S. 148 verwandt; ähnliches Vorgehen bei *Lauchs/Link/Lösel*, Die Erfassung der Deliktschwere in Evaluationsstudien zur Straftäterbehandlung: Entwicklung und Anwendung eines strafrahmenorientierten Ansatzes, *MschKrim* 2020, 300 ff.

35 Vgl. bereits *Höfer*, Sanktionskarrieren: Eine Analyse der Sanktionshärteentwicklung bei mehrfach registrierten Personen anhand von Daten der Freiburger Kohortenstudie, Freiburg i. Br. 2003, S. 36 ff., 96 ff.; sowie *Grundies*, Regionale Unterschiede in der gerichtlichen Sanktionspraxis in der Bundesrepublik Deutschland. Eine empirische Analyse, in: Hermann/Pöge (Hrsg.), *Kriminalsoziologie: Handbuch für Wissenschaft und Praxis*, Baden-Baden 2018, S. 295 ff.

Arbeiten bilden deshalb nicht das „letzte Wort“. Dass sie die empirische Sanktionsforschung in Deutschland dennoch erweitert und bereichert haben, ist jedoch unverkennbar.