

Johannes Kaspar / Katrin Höffler / Stefan Harrendorf

Datenbanken, Online-Votings und künstliche Intelligenz – Perspektiven evidenzbasierter Strafzumessung im Zeitalter von „Legal Tech“

Abstract

Der Beitrag befasst sich mit den Möglichkeiten, auf Legal Tech aufbauende Methoden evidenzbasierter Strafzumessung in das deutsche Strafrecht einzubringen. Hierzu wird zunächst geklärt, welche Formen der (zumindest partiellen) Evidenzbasierung sich für den Bereich der Strafzumessung überhaupt anbieten. Darüber hinaus werden verschiedene technische Ansätze, von Strafzumessungsdatenbanken bis hin zur automatisierten Strafzumessung mittels „schwacher“ künstlicher Intelligenz, diskutiert. Als Beispiel für eine bereits realisierte Strafzumessungsdatenbank wird dabei auf diejenige des japanischen Saiban'in-Systems verwiesen. Die Chancen und Risiken verschiedener Modelle der Legal Tech im Strafzumessungsbereich werden erörtert. Abschließend wird weiterer empirischer und normativer Forschungsbedarf konstatiert.

Schlüsselwörter: Strafzumessung, Legal Tech, künstliche Intelligenz, Strafzwecke, Evidenzbasierung, Strafzumessungsdatenbank

Abstract

The article discusses options for introducing Legal Tech methods of evidence based sentencing into German penal law. First, it is analyzed which forms of (at least partial) evidence base are feasible at all for the area of sentencing. Then different technical approaches, from sentencing databases to automated sentencing by “weak” artificial intelligence, are discussed. As an example for an already existing sentencing database the authors refer to the one used for the Japanese Saiban'in system. Chances and risks of different Legal Tech options for the area of sentencing are debated. The authors conclude that further empirical and normative research in the field is necessary.

Keywords: sentencing, Legal Tech, artificial intelligence, aims of punishment, evidence base, sentencing data base

A. Problemstellung

Der Vorgang der Strafzumessung ist für Betroffene existenziell bedeutend und belastend. Anders als bei der zunächst nur „virtuellen“, potenziell an alle gerichteten abstrakten Verbotsnorm wird bei der Bestrafung im konkreten Einzelfall ganz unmittelbar und spürbar mit staatlichem Zwang in Freiheitsgrundrechte eingegriffen. Die Entscheidung über Art und Höhe der Strafe bestimmt über das Ausmaß dieses Grundrechtseingriffs.¹ In einem Rechtsstaat bedarf sie einer möglichst klaren gesetzlichen Grundlage, die für Bestimmtheit, Gleichheit und Angemessenheit der Strafzumessung sorgt.² Das ist im geltenden Recht nur bedingt gewährleistet. Die Strafraumen sind denkbar weit ausgestaltet, die zentrale Strafzumessungsnorm in § 46 StGB ist vage gehalten und entfaltet wenig Orientierungswirkung. Die Weite und Offenheit der gesetzlichen Rahmenbedingungen dürfte einer der Gründe sein, warum in empirischen Studien auch neueren Datums immer wieder erhebliche regionale Unterschiede bei der Strafzumessung in an sich vergleichbaren Fällen festgestellt wurden,³ die sich plausibel mit abweichenden lokalen Traditionen erklären lassen. Solches gerichtliches „Strafzumessungsbrauchtum“ ist aber für sich genommen kein sachliches Differenzierungskriterium. Die Schuld wiegt in Bayern nicht schwerer als in Schleswig-Holstein, die spezialpräventiven Bedürfnisse in vergleichbaren Fällen sind in NRW nicht anders als in Berlin, die Bevölkerung in Hessen braucht nicht mehr Abschreckung als die in Sachsen. „Wege zur Strafgleichheit“⁴ auszuloten, erscheint auch deshalb wünschenswert, weil inhomogene Bestrafung ab einem gewissen Grad von der Bevölkerung als Gerechtigkeitsproblem wahrgenommen werden dürfte, was das Vertrauen in die Rechtsordnung insgesamt erschüttern kann. Empirische Studien zeigen zudem, dass es einen Zusammenhang zwischen der Wahrnehmung des Rechtssystems als gerecht oder ungerecht und der eigenen Normbefolgung gibt.⁵ Kurz gesagt: unterschiedliche Strafen für gleich gelagerte Straftaten sind nicht nur ein Gleichheits- und Gerechtigkeitsproblem, sondern auch präventiv dysfunktional.

Der überwiegend anerkannte Befund solcher Strafzumessungsungleichheit war der Ausgangspunkt des von *Kaspar* erstellten Gutachtens für den 72. Deutschen Juristentag (DJT) in Leipzig im November 2018.⁶ Die dortige kontroverse Diskussion hat erneut gezeigt, dass es einfache Lösungen in diesem Bereich nicht gibt. Und es ist eine offene Frage, inwiefern der Gesetzgeber mit seinem abstrakt-generellen Instrumentarium die Strafzumessung im Einzelfall überhaupt sinnvoll steuern kann bzw. steuern sollte. Dabei besteht die Gefahr, dass vorab kleinteilig festgelegte Vorgaben dem Einzelfall nicht gerecht werden und zu einer starren „Prokrustes-Justiz“⁷ führen. Dann

1 S. näher *Kaspar* 2018, C 90 ff.; umfassend *ders.* 2014, 619 ff.; s.a. *Harrendorf* 2020.

2 S. näher zu diesen drei Problemfeldern *Kaspar* 2019.

3 S. die Nachweise bei *Kaspar* 2018, C 18 ff.

4 *Kaspar* 2019.

5 S. nur *Robinson* 2014, 152 ff.

6 *Kaspar* 2018.

7 S. bereits *Weigend* 1988, 579, 600.

könnte es zu einer schematischen Gleichbehandlung von an sich Ungleichem kommen, was verfassungsrechtlich im Hinblick auf Art. 3 I GG nicht weniger problematisch ist als die Ungleichbehandlung von an sich Gleichem. Auch weil die Persönlichkeiten derjenigen, an die sich die Strafe richtet, so unterschiedlich sind, muss Raum für eine hinreichende Individualisierung sein, da die Strafe gem. § 46 I 2 StGB auch Aspekte der Spezialprävention und Strafempfindlichkeit zu berücksichtigen hat. All dies spricht gegen die Idee, die Strafzumessung zu schematisch an mehr oder weniger verbindlichen Tabellen oder Richtlinien (etwa im Sinne der US-amerikanischen „Federal Sentencing Guidelines“) zu orientieren, die sich tendenziell auf markante und leicht quantifizierbare Merkmale beschränken.⁸

Auf der anderen Seite ist es offensichtlich auch keine befriedigende Lösung, die Strafzumessung in das freie Ermessen des Strafgerichts zu stellen. So wurde es allerdings lange gesehen: Die Findung des Strafmaßes sei die ureigene Aufgabe des Tatgerichts,⁹ was dann konsequenterweise auch zu sehr großer Zurückhaltung bei der revisionsgerichtlichen Kontrolle der Strafzumessung führte. Viel hat sich daran bis heute nicht geändert, auch wenn nun allgemein anerkannt ist, dass die Strafzumessung eine rechtlich gebundene und auch prinzipiell überprüfbare Entscheidung ist.¹⁰ Wenn aber die Ziele und maßgeblichen Kriterien der Strafzumessung unklar sind und Gerichte (notgedrungen) mit dem sprichwörtlichen „geheimen Metermaß“¹¹ agieren bzw. einen „Griff ins Dunkle“¹² vornehmen, wird ausgerechnet eine verfassungsrechtlich hochbrisante staatliche Entscheidung weder für die direkt von ihr Betroffenen noch für die Allgemeinheit ausreichend transparent und nachvollziehbar begründet. Diese ist mit anderen Worten sehr stark von der nur schwer kontrollierbaren persönlichen Wertung der einzelnen Persönlichkeit der Richterinnen oder des Richters abhängig, was im Extremfall stets die Gefahr von Willkür beinhaltet. Und tatsächlich zeigen empirische Studien, dass neben der jeweiligen Strafzumessungstradition vor Ort auch persönliche Eigenheiten und Präferenzen der Richterinnen und Richter Einfluss auf das Strafmaß haben.¹³ Selbst die Frage, ob eine Entscheidung (hungrig) kurz vor oder (satt) nach der Mittagspause getroffen wird, scheint sich nach einer israelischen Studie auf die Härte der Entscheidung auszuwirken.¹⁴

Umso mehr stellt sich die Frage, welche Möglichkeiten bestehen, um die gerichtliche Entscheidung auf eine rationalere und nachvollziehbare Grundlage zu stellen, um die Transparenz zu erhöhen und eine effektivere (auch revisionsrechtliche) Kontrolle zu ermöglichen. Das Postulat der Rationalisierung von Entscheidungen wird in jüngerer

8 So auch das klare Abstimmungsergebnis auf dem DJT. In der Praxis existieren jedoch justizintern durchaus diverse, unverbindliche, an der lokalen Tradition orientierte Tabellen für leichtere Taten und einfache, häufig vorkommende Fälle (z.B. bei den Straßenverkehrsdelikten, s. Schäfer/Sander/van Gemmeren 2017, Rn. 1719 ff.).

9 Nachweise bei Schäfer/Sander/van Gemmeren 2017, Rn. 1506.

10 S. dazu die Nachweise bei Schäfer/Sander/van Gemmeren 2017, Rn. 1505 ff.

11 S. Dreher MDR 1961, 344.

12 V. Liszt 1905, 393.

13 Streng 1984, 131 ff.

14 Danziger u.a. PNAS 2011, 6889.

Zeit immer wieder auch mit dem Schlagwort der „Evidenzbasierung“ umschrieben. So heißt es etwa im Koalitionsvertrag der aktuellen Regierungskoalition auf S. 133: „Wir treten für eine evidenzbasierte Kriminalpolitik ein. Wir wollen, dass kriminologische Evidenzen sowohl bei der Erarbeitung von Gesetzentwürfen als auch bei deren Evaluation berücksichtigt werden.“ Hier wird auf allgemeine kriminalpolitische Maßnahmen des Gesetzgebers Bezug genommen,¹⁵ von der Strafzumessung ist nicht ausdrücklich die Rede. Wenn man aber davon ausgeht, dass auch sie (einzelfallbezogene) Kriminalpolitik darstellt, dass also durch die Art und Höhe der Strafe in sinnvoller und möglichst „wirksamer“ Weise auf Kriminalität reagiert werden soll,¹⁶ dann müsste die Forderung nach Evidenzbasierung prima facie auch dort gelten. Gerade der für alle Grundrechtseingriffe geltende Verhältnismäßigkeitsgrundsatz¹⁷ gebietet ein rationales Vorgehen des Staates, was man auch mit der Forderung nach einem möglichst evidenzbasierten Vorgehen umschreiben kann.¹⁸ Es gilt, den konkret betroffenen Personen anhand eines legitimen Zwecks darlegen zu können, warum man ihnen eine Strafe dieser Art und Höhe auferlegt.

Im Folgenden soll es daher um die Frage gehen, ob sich auch im Bereich der Strafzumessung eine stärkere „Evidenzbasierung“ anbietet, welche Probleme sich dabei stellen und welche Rolle die Möglichkeiten der modernen Datenverarbeitung spielen könnten. Das potentielle Spektrum EDV-gestützter Strafzumessung ist breit und reicht von der lediglich unterstützten vollständig eigenen Entscheidung des Gerichts bis hin zur Utopie (oder Dystopie) einer ganz eigenständig durch den Einsatz von künstlicher Intelligenz (KI) vorgenommenen Strafzumessung. Die Perspektiven des Einsatzes von KI im Rechtssystem werden zunehmend diskutiert, wobei sich auch aufgrund des rasanten technologischen Fortschritts derzeit nur schwer und sicherlich unvollständig abschätzen lässt, wie sinnvolle Anwendungsmöglichkeiten aussehen könnten und welche Probleme und Gefahren dabei bestehen. Bereits heute ist „Legal Tech“ in einem weniger ambitionierten Sinne im Einsatz, z.B. in Japan im Sinne computergestützter Strafzumessung mit Hilfe von Datenbanken.

B. Begriffsklärungen

Bevor diesen Fragen näher nachgegangen werden kann, müssen einige begriffliche Klärungen vorgenommen werden.

15 Zu den Schwierigkeiten der „Evidenzbasierung“ in diesem Bereich s. *Kaspar KriPoZ* 2020, 14 ff. und *Höffler*, 2018, 225 ff.

16 Das ist natürlich eine anfechtbare Prämisse, die mit der Diskussion der Strafzwecke zusammenhängt, dazu sogleich im Text.

17 Zu dessen Bedeutung im materiellen Strafrecht s. umfassend *Kaspar* 2014.

18 Zur Frage einer „Evidence Based“ Kriminalpolitik“ *Höffler* 2018, 225 ff., s. dort auch 234 zum Verhältnis von „rational“ zu „evidenzbasiert“ und wohlfahrtsökonomischen Gesichtspunkten.

I. Evidenzbasierung

Mit der Forderung nach „Evidenzbasierung“ ist hier gemeint, dass Maßnahmen auf der Grundlage des aktuellen Stands der wissenschaftlichen Erkenntnis über (objektive) tatsächliche Zusammenhänge getroffen werden.¹⁹ Man kennt diese Ausdrucksweise vor allem aus dem medizinischen Bereich,²⁰ wo diese auf eine wissenschaftliche Fundierung der angenommenen Wirksamkeit ärztlicher Behandlungen abzielt. Gerade dieses Beispiel zeigt sehr gut, dass es um eine empirische und nicht um eine normative Perspektive geht. Das ist gerade im juristischen Kontext wichtig, wo in der Regel die letztere im Vordergrund steht. Vielleicht auch deshalb wird „Evidenz“ hier oft in einem abweichenden Sinn verstanden, nämlich als „ohne weiteren Begründungsaufwand offensichtlich“ oder dem Wortsinn entsprechend „dem Augenschein“ nach.²¹ Das kann zu Missverständnissen führen und hat manche dazu bewogen, andere Bezeichnungen vorzuschlagen, also etwa „nachweisbasiert“ statt „evidenzbasiert“.²² Hier soll an letzterem Terminus²³ festgehalten werden, da er nicht nur in der Politik zunehmend verwendet wird, sondern auch im strafrechtswissenschaftlichen und kriminalpolitischen Diskurs fest etabliert ist. Insofern kann auf die US-amerikanische Strömung der „evidence-based crime prevention“ verwiesen werden,²⁴ die in Deutschland etwa in Form des „Düsseldorfer Gutachtens“ aufgegriffen²⁵ und jüngst auch in Form eines wissenschaftlich fundierten „Leitfadens für Politik und Praxis“ umfassend vertieft wurde.²⁶

II. Legal Tech

Mit Legal Tech (Kurzform von Legal Technology)²⁷ ist im weitesten Sinne Informatikstechnik gemeint, die das juristische Arbeiten in irgendeiner Form unterstützt. Gegenüber der eher wissenschaftlich orientierten Rechtsinformatik ist Legal Tech im Wesentlichen anwendungsbezogen und dient der effizienten Lösung konkreter Probleme. Legal-Tech-Anwendungen finden sich vor allem im privatrechtlichen Bereich, etablieren sich jedoch langsam in öffentlich-rechtlichen Rechtsbeziehungen und gelangen darüber auch ins Strafrecht.²⁸ Das Komplexitätsniveau derartiger Anwendungen ist facettenreich: Sie reicht von Software zur Arbeits- und Büroorganisation (Legal Tech 1.0) über die Teilautomatisierung repetitiver juristischer Arbeitsschritte (Legal Tech 2.0) bis

19 Zum Folgenden s. auch *Hamann* 2014, 2 f.

20 Nachweise bei *Hamann* 2014, 3 f.

21 S. dazu auch *Höffler* 2018, 234 Fn. 47.

22 So etwa *Coester* 2018, 37 ff.

23 I.S.e. empirischen Evidenz.

24 S. nur *Sherman* u.a. 2002.

25 *Rössner* u.a.; s.a. *Coester* 2018, 37 ff.

26 *Walsh* u.a. 2018.

27 S. dazu auch *Kilian*, NJW 2017, 3043, 3048 ff.

28 Zum Einsatz von Legal Tech im Bereich der Strafzumessung s. *Rostalski/Völkening* KriPoZ 2019, 265 ff.; zu Predictive Policing *Sommerer* 2020; *Singelstein* NSTz 2018, 1.

hin zur algorithmisch-automatisierten digitalen Rechtsfindung ohne menschliche Mitwirkung durch künstliche Intelligenz (Legal Tech 3.0).²⁹ Während der Einsatz von Datenbanken nach japanischem Vorbild bspw. eher auf der Stufe von Legal Tech 1.0 anzusiedeln wäre, gingen (teil-)automatisierte Decision-Support-Systeme in der Strafzumessung schon über die Mindestanforderungen für Legal Tech 2.0 hinaus, freilich ohne die Freiheit von menschlicher Mitwirkung der Stufe 3.0 zu erreichen. Legal-Tech-Anwendungen scheinen dabei im Bereich der Strafgerichtsbarkeit Effizienzpotentiale mit sich zu bringen,³⁰ werfen aber gerade dort natürlich zentrale normative Fragen auf, die im Folgenden aufgegriffen werden.

III. Künstliche Intelligenz

Vor allem bei den fortgeschrittenen Entwicklungsstufen von Legal Tech kann KI zum Einsatz kommen. Bei KI handelt es sich um ein Teilgebiet der Informatik. Ziel ist es, menschliches Entscheidungsverhalten durch Algorithmen zu imitieren bzw. allgemein intelligentes automatisiertes Verhalten zu ermöglichen.³¹ Es wird zwischen „starker“ und „schwacher“ KI unterschieden: Mit starker KI ist die – bisher nicht verwirklichte – Idee verbunden, dem menschlichen Denken ebenbürtige oder es transzendierende intelligente Agenten zu schaffen. Schwache KI umfasst demgegenüber problemspezifische Anwendungen der unterschiedlichen KI-Teilaspekte, wie bspw. Spracherkennung.³² Zentral für jeden KI-Agenten ist das Lernen auf Basis von wahrgenommenem Daten-Input, etwa Bildern, Texten oder Sprache,³³ aus der jeweiligen Test- bzw. Anwendungsumgebung. Auf Grundlage des wahrgenommenen Inputs versucht der Agent durch Wiederholung der ihm gestellten Aufgabe die vorgegebenen Leistungsmaße zu erfüllen und optimiert so – bezogen auf das festgelegte Ziel – seine Leistungsfähigkeit.³⁴ Vor allem aufgrund ihrer „regulatorischen Steuerungsmacht“ stehen die verschiedenen Spielarten der KI in einem Konkurrenzverhältnis zum traditionellen Steuerungsanspruch des Rechts und im Spannungsverhältnis zu den Werten der europäischen Rechtstradition, so dass in sensiblen Bereichen wie dem der Strafzumessung eine tiefgehende Befassung mit Technologie, sozialen Nebenfolgen und effektiven regulatorischen Rahmenbedingungen geboten ist.³⁵

29 Vgl. *Groh* 2019b.

30 So etwa *Kleinberg* u.a., *The Quarterly Journal of Economics*, 133 (2018), 237 ff., für die Entscheidung von US-Richtern, ob Untersuchungshaft wegen Wiederholungsgefahr angeordnet werden soll.

31 Siehe dazu den Überblick bei *Russel/Norvig* 2016, 2. Zu den definitorischen Unschärfen s. *Herberger* NJW 2018, 2825 ff.

32 *Groh* 2019a; für weitere Bsp. siehe *Russel/Norvig* 2016, 28 f.

33 *Groh* 2019a.

34 *Russel/Norvig* 2016, 59.

35 *Martini* 2019, 9 ff.

C. Probleme „evidenzbasierter“ Strafzumessung

Die entscheidende Frage ist nun, ob es „evidenzbasierte“ Strafzumessung überhaupt geben kann, wenn man mit „Evidenz“ tatsächliche Zusammenhänge meint, die mit den Mitteln der empirischen Forschung erfasst werden und dann einer späteren Wirkungskontrolle zugrunde gelegt werden können. Dabei geht es noch nicht einmal in erster Linie um die Frage, wie wissenschaftliche Erkenntnisse in den richterlichen Entscheidungsprozess einfließen können: Hier bedarf es ganz offensichtlich der praxistauglichen Aufbereitung dieser Erkenntnisse, bei der möglicherweise „Legal Tech“ eine enorme Hilfe sein kann, dazu später mehr. Davor muss aber die ganz grundsätzliche Frage geklärt werden, ob die Strafzumessung in ihrer derzeitigen rechtlichen und praktischen Ausgestaltung überhaupt hinreichend anschlussfähig ist für eine „Evidenzbasierung“ und eine darauf gestützte Wirkungskontrolle. Wenn man die gängigen Strafzwecke betrachtet, ist dies offensichtlich nicht so einfach.

I. Schuldausgleich

Das betrifft vor allem die Strafzumessung im engeren Sinn, die die Höhe der Strafe für eine bestimmte Straftat betrifft. Strafe soll nach der Grundlagenformel in § 46 I 1 StGB in erster Linie anhand der durch die Tat verwirklichten „Schuld“ bemessen werden. Legt man diesen retributiven Strafzweck zugrunde, fehlt es an einer mit der Strafe verbundenen Wirkerwartung und damit an der Grundvoraussetzung einer evaluierenden Kontrolle der Bestrafung im Hinblick auf die Erreichung eines vorab definierten empirischen Zwecks. Dies macht die Retribution auch verfassungsrechtlich problematisch.³⁶ Die Grundlage der gerichtlichen Schuldbewertung sind zwar auch faktische Elemente (in der Praxis bspw. die Zahl der Vorstrafen),³⁷ und insofern kann die Basis der Entscheidung sachlich falsch oder richtig sein. Für die Wertungsentscheidung selbst, also die „Umrechnung“ bestimmter angenommener Umstände in ein konkretes Schuld- und damit letztlich Strafmaß gilt das aber für das in Deutschland bisher zugrunde gelegte, relativ weite Verständnis von Strafzumessungsschuld nicht. Ob das Strafmaß „richtig“ in diesem Sinne ist, wäre dann eine Frage, die man prinzipiell nicht in objektivierbarer Weise beantworten könnte.

Anderes dürfte jedoch gelten für eine Vergeltungsstrafe, die allein nach dem Maßstab der Tatproportionalität verhängt würde.³⁸ Hier würden nämlich alle potentiellen Deliktsverwirklichungen auf einer durch das Strafgesetz fixierten Skala von der denk-

36 S. dazu näher *Kaspar* 2014, 134 ff.; *Harrendorf* 2020; s.a. *Harrendorf* 2018, 97, 101.

37 Gegen eine strafscharfende Berücksichtigung von Vorstrafen spricht allerdings auch aus Vergeltungsperspektive einiges; siehe näher und rechtsvergleichend *Roberts/Harrendorf* 2020, 261 ff.; zu einem konkreten Vorschlag zur Begrenzung der Strafzumessungsrelevanz von Vorstrafen s. *Kaspar* 2018, C 104 ff.

38 Siehe hier nur *Hirsch/Jareborg* 1991; *Hörnle* 1999; nähere Auseinandersetzung mit derartigen Ansätzen und generalpräventive Umformulierung bei *Harrendorf* 2020; s. dazu auch *Kaspar* 2018, C 38 ff.

bar leichtesten bis zur denkbar schwersten Deliktsverwirklichung angeordnet werden können. Berücksichtigungsfähig wären nur das Handlungs- und Erfolgsunrecht und (lediglich mildernd) auch Ausprägungen der Strafbegründungsschuld, die hinter dem gesetzlichen Regelfall voller Schuldfähigkeit zumindest graduell zurückbleiben.³⁹ Unter der Prämisse, dass sich alle (inneren und äußeren) Schwermerkmale einer Tat empirisch operationalisieren lassen können, wäre durchaus ein umfassender, individualisierter Deliktsschwereindex denkbar, der in Strafmaße automatisiert umrechenbar wäre.

Selbstverständlich wäre Ansatzpunkt maschinellen Lernens bei Zugrundelegung einer solchen Theorie weiterhin nicht der „Erfolg“ der Sanktion im Sinne einer präventiven Zweckerfüllung: Ein „Lernen“ am Misserfolg (sei es durch Menschen oder durch Maschinen) ist auch hier nur schwer vorstellbar. Und genau das macht den retributiven Strafzweck ja für viele so attraktiv: Auf einen (etwa präventiven) Nutzen der Strafe kommt es nicht an, so dass jegliche kriminologische Erkenntnis über Wirksamkeit oder Unwirksamkeit von Strafe außer Acht gelassen werden kann. Dies spricht aus unserer Sicht auch gegen eine vergeltende Tatproportionalitätslehre.

II. Spezial- und Generalprävention

Wie sieht es nun mit den gesetzlich ebenfalls vorgesehenen präventiven Zielen der Strafe aus? Bieten sie mehr Ansatzpunkte für Evidenzbasierung? Zwar wird in § 46 I 2 StGB auf den (auch) spezialpräventiven Aspekt der „Wirkungen, die von der Strafe für das künftige Leben des Täters in der Gesellschaft zu erwarten sind“ abgestellt, der einer Evidenzbasierung prinzipiell zugänglich ist. Aber dieser Aspekt soll sich nach h.M. nur im Rahmen des „Schuldangemessenen“ entfalten und hat daher nur begrenzten Einfluss.⁴⁰ Auch ist die Norm so offen formuliert, dass sie alle möglichen Effekte der Strafe umfasst, von der (erhofften) Abschreckungswirkung bis zur (näherliegenden) großen Belastung für das spätere Leben in der Gesellschaft. Das ist schwer zu operationalisieren und erschwert eine differenzierte Bewertung einzelner Strafzumessungsniveaus.

Dasselbe gilt für die in § 46 StGB nicht explizit erwähnte negative Generalprävention, deren Berücksichtigung innerhalb des Schuldrahmens aber nach der Rechtsprechung dennoch zulässig sein soll.⁴¹ Sie wird, wenn überhaupt, als schärfender Faktor herangezogen. Das wiederum setzt voraus, dass eine Strafschärfung im Einzelfall messbar abschreckende Wirkungen entfaltet – eine mehr als zweifelhafte Annahme, wenn man ernsthaft „evidenzbasiert“ vorgeht und die Erkenntnisse der Generalpräventionsforschung berücksichtigt, nach denen es für einen Abschreckungseffekt praktisch ausschließlich auf die subjektiv antizipierte Entdeckungs- und Sanktionierungswahr-

³⁹ Hörnle 1999, 151, 328 f.

⁴⁰ Sog. „Spielraumtheorie“, z.B. BGHSt 7, 28; 24, 132; *Miebach/Maier* 2016, § 46 Rn. 47.

⁴¹ St. Rspr., s. nur BGHSt 28, 318, 326 f.; BGH NStZ 1992, 275; zusammenfassend und m.w.N. *Schäfer/Sander/van Gemmeren* 2017, Rn. 839 ff.

scheinlichkeit und nicht auf die Höhe der drohenden Strafe ankommt.⁴² Auch ist zu berücksichtigen, dass weder die Grenzpunkte des Schuldrahmens noch das Ausmaß spezial- oder generalpräventiver Modifikationen der reinen Schuldstrafe vom Gericht im Urteil benannt werden, so dass diese Aspekte letztlich untrennbar ineinander übergehen. Eine darauf gestützte evidenzbasierte Kontrolle ist also auch deshalb schwer möglich. Soweit es schließlich um den Gesichtspunkt positiver Generalprävention geht, wird dessen Berücksichtigung zwar überwiegend für zulässig gehalten; häufig wird aber recht kurzschlüssig die schuldangemessene mit der positiv generalpräventiv erforderlichen Strafe gleichgesetzt,⁴³ so dass jedenfalls eine separate empirische Berücksichtigung positiver Generalprävention im Kontext konkreter Strafzumessung bisher nicht zu vermuten ist.

Bei den Entscheidungen der Strafzumessung im weiteren Sinne (z.B. bei der Entscheidung über die Straftat Geld- oder Freiheitsstrafe gem. § 47 StGB oder bei der Frage der Aussetzung der Freiheitsstrafe zur Bewährung gem. § 56 StGB) gestaltet sich das Bild etwas anders:⁴⁴ Hier soll es schon nach dem Willen des Gesetzgebers explizit auch darum gehen, dass die verurteilte Person selbst keine Straftaten mehr begeht. Zumindest dieses Zielkriterium ist klar definiert und kann prinzipiell empirisch erforscht werden. Allerdings ist es bekanntlich äußerst schwer, in dieser Hinsicht aussagekräftige Ergebnisse zu erzielen. Insbesondere lässt sich eine kausale Wirkung der Sanktion auf die Legalbewährung nur messen, wenn mit möglichst homogenen Kontroll- und Vergleichsgruppen gearbeitet wird, die idealerweise durch zufällige Verteilung gebildet werden – gerade im Bereich strafrechtlicher Sanktionierung ist das regelmäßig nicht zu realisieren. Prinzipiell sind hier aber Forschungsergebnisse möglich und existieren auch bereits, die man der richterlichen Legalprognose zugrunde legen könnte. Bislang wird das in Deutschland zwar nicht ausreichend praktiziert, hier besteht aber durchaus ein ausbaufähiger Bereich für anwendungsbezogene Forschung, die dann auch wichtige Implikationen für die Praxis entfalten könnte. Dass dabei allerdings noch einiges zu tun ist, zeigen Erfahrungen aus dem Ausland, wo Legal Tech im Sinne von automatisierten Risikoprognosetools teils bereits großflächig eingesetzt wird, Evaluierungen aber zeigen, dass die entscheidenden Instrumente oft noch nicht ausgereift sind.⁴⁵

Das Problem der nur schwachen empirischen Basis der justiziellen Entscheidungspraxis zeigt sich auch (und vielleicht noch stärker) beim (vorwiegend positiv-)general-

42 S. nur die Meta-Analyse von *Spigath* 2013. Zur Vereinbarkeit mit den Grundannahmen des rational-choice-Ansatzes s. *Harrendorf/Geng* 2019, S. 189 f.

43 Z.B. bei *Dölling*, ZStW 102 (1990), 1, 16; *Müller-Dietz* 1985, 813, 823 f.; ausführlich zum Problem *Harrendorf* 2020.

44 S. dazu umfassend *Schulte* 2016.

45 Zum in den USA vielfach eingesetzten COMPAS-Programm zur Rückfallprognose siehe z.B. *Dressel/Farid*, Science Advances 4 (2018), die zeigen konnten, dass dieses Programm, obwohl es immerhin 137 verschiedene Variablen integriert, keine besseren Prognosen anstellt als Personen ohne oder mit nur wenig strafrechtlich-kriminologischen Vorkenntnissen. Bereits zuvor zeigte eine Studie, dass dem Programm eine Benachteiligung von afroamerikanischen Menschen eingeschrieben ist; die Rate der false positives war hier doppelt so hoch, vgl. *Angwin* u.a. ProPublica (2016). Dazu auch *Höffler* 2019, 60.

präventiven⁴⁶ Aspekt der „Verteidigung der Rechtsordnung“, der in den §§ 44, 47, 56 und 59 StGB erwähnt wird.⁴⁷ Er stellt dort jeweils eine Art „Notbremse“ dar für Entscheidungen, die unter generalpräventiven Aspekten vermeintlich zu milde sein könnten – ein erneuter Beleg für die zweifelhafte und an sich schon lange widerlegte Grundannahme des Gesetzgebers, dass die härtere Strafe generalpräventiv stets besser „wirkt“. Wenn nun das Gericht in einem Fall beispielsweise annimmt, dass zwingend zur Freiheitsstrafe gegriffen werden müsse, weil eine Geldstrafe zur „Verteidigung der Rechtsordnung“ nicht genüge, so wird von ihm bezeichnenderweise nicht verlangt, dass es diese Annahme auf empirische Fakten stützt. Man spricht zwar von „Generalprävention“, lässt aber letztlich die bloße Behauptung eines generalpräventiven „Bedürfnisses“ im konkreten Fall genügen, die nach der Rechtsprechung nicht mit möglicherweise gegenläufigen empirischen Forschungsergebnissen in Frage gestellt werden kann.⁴⁸ Hier zeigt sich erneut „in a nutshell“ das Problem der Strafzumessung insgesamt: Die richterliche Entscheidung wird im Ergebnis ganz maßgeblich auf eine Wertung gestützt, der keine überprüfbaren tatsächlichen Zusammenhänge oder auch geprüfte Wirkungsannahmen zugrunde liegen.

III. Weitere Bezugspunkte

Als Bezugspunkt der „Evidenzbasierung“ kommen neben unmittelbar strafzweckbezogenen Wirkungszusammenhängen zwei weitere prinzipiell empirisch messbare Umstände in Betracht: Die bisherige Strafzumessungspraxis in vergleichbaren Fällen sowie Erkenntnisse über Strafbedürfnisse und Gerechtigkeitsvorstellungen der Bevölkerung.

1. Bisher übliches Strafniveau

Zunächst wäre es denkbar, die Strafzumessungsentscheidung am bisherigen durchschnittlichen bundesweiten Strafniveau für vergleichbare Fälle zu orientieren, wie es immer wieder gefordert wurde,⁴⁹ in jüngerer Zeit auch unter dem Oberbegriff der „komparativen Strafzumessung“.⁵⁰ Ob sich das rechtfertigen lässt, hängt vom straftheoretischen Standpunkt ab. Wer auf Spezialprävention setzt, müsste auf einer starken Individualisierung der Strafzumessung bestehen, bei der ein vergleichender Seitenblick auf das Strafniveau in ähnlichen Fällen wenig Aussagekraft besitzt. Auch wer die Retribution als Strafzweck befürwortet, müsste mit einem vergleichenden Maßstab ein gewisses Problem haben, denn ob die bisherige Strafpraxis das „richtige“, schuldangemessene Niveau abbildet, wäre noch zu begründen. Immerhin könnten die bisher zur

46 Näher *Harrendorf*, Criminal Law Forum 28 (2017), 501, 512 ff. m.w.N.

47 Zu dessen empirischem Gehalt s. *Eierle* 2019, 163 ff.

48 Ausdrücklich so: OLG Celle JR 1980, 256, 257; BayObLG NJW 1978, 1337; näher *Harrendorf* 2020.

49 *Schöch* 1973, 76 f.; *Streng* 1984, 299 ff.

50 Vgl. *Maurer* 2005.

Entscheidung berufenen Gerichte ja allesamt daneben gelegen haben, während man selbst nun das „wahre“ Schuldmaß in dieser Konstellation entdeckt hat (was nur ein bisschen polemisch ist, wenn man sich die Unbestimmtheit des Schuld-Kriteriums im Strafzumessungskontext nochmals vor Augen führt). Zumindest könnte man pragmatisch sagen, dass eine Vielzahl von Personen in der Vergangenheit in vergleichbaren Fällen durchschnittlich zu diesem Strafmaß gekommen ist, so dass man annehmen kann, dass sie damit eine (auch für die Zukunft) akzeptable Schuldwertung getroffen haben – richterliche „Schwarmintelligenz“ sozusagen. Noch naheliegender ist ein solcher Maßstab, wenn man im Sinne der positiven Generalprävention mehr auf die faktisch im Großen und Ganzen in der Vergangenheit eingetretene friedensstiftende Wirkung der Strafzumessungspraxis abstellt, die auch in Zukunft zu erwarten ist. Mit dem Umstand, dass dieses Niveau regional z.T. sehr unterschiedlich ausfällt, könnte im Sinne der Verhältnismäßigkeit so verfahren werden, dass im Zweifel das niedrigere Niveau angestrebt wird, da es das „generalpräventive Minimum“ zu finden gilt. Zudem könnte für unter spezialpräventiven Gesichtspunkten deutlich vom Regelfall abweichende Konstellationen ein Spielraum vorgesehen bleiben.

2. Strafbedürfnisse der Bevölkerung

Zumindest wenn man einer positiv generalpräventiven Theorie folgt, liegt es nahe, noch einen Schritt weiter zu gehen und nicht (allein) auf die bisherige Strafpraxis als Indiz für Friedensstiftung abzustellen, sondern direkt auf das zur Friedensstiftung aus Sicht der Allgemeinheit Erforderliche. Damit kommen Strafbedürfnisse und Gerechtigkeitsvorstellungen der Bevölkerung als Gegenstand der empirischen Forschung in Betracht. Bei allen methodischen Schwierigkeiten, die hier bestehen, dürfte es möglich sein, in repräsentativen Umfragen valide Ergebnisse darüber zu erzielen, was die Bevölkerung in bestimmten, genauer beschriebenen Konstellationen für eine aus ihrer Sicht „angemessene“ Strafe hält. Ob (und wenn ja: wie) man diesen empirischen Umstand dann der Strafzumessungsentscheidung wirklich zugrunde legen sollte, ist hoch umstritten.⁵¹ Jedenfalls kann hier (wie beim Bezugspunkt des bisher üblichen Strafmaßes) nur in einem eingeschränkten Sinne von „Evidenzbasierung“ gesprochen werden: Man stützt (immerhin) die Entscheidung auf empirisch messbare tatsächliche Umstände – nicht dagegen geht es dabei um Wirkungszusammenhänge der Strafe. Überprüfbar ist mit anderen Worten nur, ob das Urteil der bisherigen Strafpraxis oder den Vorstellungen der Allgemeinheit entspricht – ob es geeignet und erforderlich ist, die gewünschten (generalpräventiven) Wirkungen zu erzielen, ist damit noch nicht gesagt.

Dabei verbietet es sich aus verschiedenen Gründen, gesellschaftliche Strafbedürfnisse bei der Strafzumessung jedenfalls schärfend zu berücksichtigen, nicht zuletzt aufgrund der richterlichen Unabhängigkeit und des Primats des demokratisch legitimierten Strafgesetzgebers. Zudem ist aus methodischer Sicht der Einwand zu erheben, dass es eher auf ein „public judgement“, also ein abgewogenes öffentliches Urteil, als bloß

⁵¹ Vgl. nur die Beiträge in *Kaspar/Walter* 2019.

auf „raw public opinion“, also die ad hoc aus dem Bauch heraus geäußerte öffentliche Meinung ankommen sollte.⁵² Ein „public judgement“ wäre allein mit Meinungsumfragen nicht erzielbar, denkbar wären aber „deliberative polls“, bei denen Personen aus der allgemeinen Bevölkerung vor ihrer Entscheidung zunächst – möglichst wertneutral – über die rechtlichen Rahmenbedingungen und kriminologischen Wirkungszusammenhänge informiert werden, bevor sie eine Entscheidung über das Strafmaß treffen.⁵³

Sollte sich aber zeigen lassen, dass die Bevölkerung in bestimmten Bereichen *weniger punitiv* ist als die Strafjustiz, kann dem durchaus durch eine (vom Gesetzgeber initiierte oder auch nur vom Gericht innerhalb der geltenden Strafraumen vorgenommene) Absenkung des Strafniveaus Rechnung getragen werden. Auch insofern könnte die EDV in Zukunft Hilfe leisten, wenn es dem Gericht möglich ist, empirische Daten zu Straferwartungen in vergleichbaren Konstellationen durch wenige Mausklicks abzurufen, um zumindest einen „Ankerwert“⁵⁴ für seine weiteren Überlegungen zu schaffen. Und warum sollte man in Zukunft nicht den Gedanken der Laienbeteiligung im Justizsystem radikal weiterdenken: Wenn Schöffinnen und Schöffen das „Rechtsgefühl“ der Allgemeinheit in die konkrete Aburteilung einbringen sollen,⁵⁵ warum soll dann nicht irgendwann ein Panel von repräsentativ ausgewählten „E-Schöffinnen“ und „E-Schöffen“, die eine juristische und kriminologische Schulung durchlaufen haben, zu einem vom Gericht aufbereiteten Fall eine Strafmaßempfehlung per Online-Voting abgeben?

D. Formen computergestützter Strafzumessung

I. Allgemeine Informationssysteme

Bereits jetzt stehen den in der Justiz tätigen Personen allgemeine Informationssysteme zur Verfügung, um ihre Entscheidungen zu untermauern, vor allem Datenbanken wie Beck Online, Jurion oder Juris, die eine Suche nach vergleichbaren Urteilen ermöglichen. Legal Tech kommt hier also zum Einsatz, aber nur in der 1.0-Variante, im Wege eines „Nachschlagens“ von einschlägigen Entscheidungen und Aufsätzen. Auf diese Weise ermittelte Informationen fließen dann (genauso wie andere, nicht-elektronische Informationsquellen) in die Strafzumessungsentscheidung ein, ohne dass dafür ein formaler Rahmen besteht. Zur Frage, ob und wie vertieft sich Richterinnen und Richter auf diese Weise informieren, existieren keine rechtlichen Vorgaben. Das hat den Vorteil, dass Gewaltenteilung und richterliche Unabhängigkeit gem. Art. 97 GG vollständig gewahrt bleiben. Auf der anderen Seite steht die Offenheit und Informalität der

52 Instruktiv und ausführlich *Green* 2008, 241 ff.; s.a. *Harrendorf* 2020; *Roberts/Harrendorf* 2020, 276.

53 *Luskin* u.a., *British Journal of Political Science* 32 (2002), 455 ff.; *Cochran/Chamlin*, *Journal of Criminal Justice* 33 (2005), 573 ff.; *Indermaur* u.a., *Punishment & Society* 14 (2012), 147 ff.

54 Allerdings können Ankerwerte die Unabhängigkeit und Rationalität der richterlichen Entscheidung auch gefährden, s. nur *Englich* u.a., *Personality and Social Psychology Bulletin* 32 (2006), 188 ff.

55 Vgl. nur *Schmidt* 2019, 33.

Vorgehensweise der anzustrebenden Rationalisierung und Homogenisierung der Strafzumessung entgegen. Das gilt letztlich auch für die (wünschenswerte) Verbesserung der statistischen Erfassung von Strafzumessungsentscheidungen.⁵⁶

II. Spezifische Strafzumessungsdatenbanken

Insofern könnte die Schaffung einer spezifischen Datenbank ein Fortschritt sein, die möglichst vollständige detaillierte Daten über die Strafzumessungspraxis enthält. Den Gerichten wäre es dann nicht nur möglich, anhand bestimmter Suchkriterien eine Vielzahl von vergleichbaren Fällen zu identifizieren. Das System könnte vielmehr zugleich eine statistische Auswertung der jeweils verhängten Strafmaße liefern, die dann (etwa als Anhaltspunkt für den Einstieg in den Strafraumen) der Entscheidung zugrunde gelegt werden könnte. Ein großer Vorteil könnte sich auch im Hinblick auf die revisionsrechtliche Kontrolle durch den BGH zeigen, werden doch Strafzumessungsentscheidungen auch bisher schon gelegentlich aufgehoben, wenn sie sich vom „üblichen Strafniveau“ in vergleichbaren Fällen zu sehr entfernen.⁵⁷ Letztlich entscheidend ist aber auch hier bisher ein subjektiver Eindruck der BGH-Richterinnen und -Richter in Bezug auf die Fälle, die sie kraft Zuständigkeit zu Gesicht bekommen – eine kleine, nicht repräsentative Auswahl. Insofern wäre es ein Fortschritt, wenn sich der BGH über das tatsächliche durchschnittliche Strafniveau in vergleichbaren Fällen informieren könnte. Im Anschluss an einen Vorschlag des Gutachters *Kaspar*⁵⁸ hat sich eine Mehrheit der Abstimmenden auf dem DJT 2018 für die Einführung einer solchen Datenbank ausgesprochen. Dabei wurde ausdrücklich auf das Beispiel Japans Bezug genommen.

Zum Hintergrund der 2009 eingeführten Strafzumessungsdatenbank in Japan muss man wissen, dass auch dort sehr weite Strafraumen (mit der Todesstrafe als Endpunkt) und große Ermessensspielräume der Gerichte bestehen. Dennoch scheint die Strafzumessungspraxis in Japan sehr viel homogener zu sein als in Deutschland – man spricht nicht ohne Stolz von einem etablierten und landesweit einheitlichen „Strafzumessungskurs“. Einer der Gründe ist offenbar, dass die Ausbildung der Richterinnen und Richter zentral erfolgt und diese dann häufig an unterschiedliche Orte versetzt werden, so dass sich keine spezifischen regionalen Niveaus entwickeln können.⁵⁹ Auch hat das Streben nach Gleichmäßigkeit sowie der Respekt vor Traditionen und Hierarchien in Japan generell einen hohen Stellenwert.

Obwohl die Orientierung am Maßstab vergleichbarer Urteile also offenbar in der Vergangenheit schon recht gut funktioniert hat, wurde 2009 eine Strafzumessungsdatenbank eingeführt. Dies hatte den speziellen Hintergrund, dass gleichzeitig eine Beteiligung von Laienrichterinnen und -richtern (japanisch: Saiban'in) eingeführt wurde, die in Fällen von Schwerdelinquenz, bei der auch die Todesstrafe im Raum steht, über die

56 S. dazu *Heinz* (in diesem Heft).

57 S. *Kaspar* 2018, C 101 f. m.w.N.

58 *Kaspar* 2018, C 120.

59 Vgl. dazu und zum Folgenden *Hieda* 2019; s. auch *Schmidt* 2019, 156.

Schuldfrage und die Strafzumessung mitentscheiden sollen. Man wollte mit dieser Maßnahme das Vertrauen der Bevölkerung in das Justizsystem stärken und deren Rechtsgefühl in die Strafzumessung einfließen lassen.⁶⁰ Zugleich fürchtete man aber offenbar, dass die Laienbeteiligung das homogene Strafniveau zu sehr durcheinander bringen könnte.⁶¹ Mit der Datenbank sollte daher gerade den Saiban'in das bisher übliche Strafniveau vor Augen geführt werden.

Zu ihrer Erstellung wurden zunächst von der Kriminalabteilung des japanischen Obersten Gerichtshofs (OGH) die Ansichten von Strafrichterinnen und -richtern in ganz Japan über relevante Strafzumessungsfaktoren ermittelt. Dann wurde auf dieser Grundlage ein privates Unternehmen mit der Erstellung der Datenbank beauftragt. Seitdem wird jedes neue Urteil in einem Saiban'in-Verfahren in der Datenbank erfasst; laut *Kato* waren dies von Mai 2009 bis Juni 2019 ca. 12.000 Urteile (etwa 2 % aller Strafsachen).⁶²

Erfasst werden spezifische Tatumstände (bspw. Taterfolg, Verhalten des Täters oder der Täterin, Beweggründe) sowie allgemeine Umstände (z.B. Reue, Wiedergutmachung, Vorstrafen), die von dem zuständigen Gericht in das System eingegeben werden. Die Bediensteten der Kriminalabteilung des OGH, die die Datenbank verwalten, überprüfen die Dateneingabe.

Nutzt das Gericht die Datenbank zur Suche vergleichbarer Fälle, erhält es ein Balkendiagramm mit der Verteilung der jeweiligen Strafmaße. Daneben zeigt das System eine Übersichtstabelle mit den einzelnen Fällen und ihren Merkmalen an, u.a. die ausgeworfene Strafe, den Strafantrag durch die Staatsanwaltschaft und eine kurze Zusammenfassung des Sachverhalts mit einem Umfang von bis zu 100 Wörtern. Der Datenbank kann auch entnommen werden, welche Umstände im jeweiligen Fall zur Annahme eines minder oder besonders schweren Falles geführt haben. Entscheidend für die Suche nach vergleichbaren Fällen ist auch die Einordnung in den jeweiligen „Typ“ der Deliktsbegehung. So wird beim Totschlag etwa unterschieden zwischen der Tatbegehung aus finanziellen Streitigkeiten, aus zufälligem Streit oder aus Erschöpfung über die langfristige Pflege kranker Verwandter.

Genutzt werden die Daten zunächst von den Strafgerichten in Saiban'in-Verfahren. Sie dienen ihnen dazu, den Spielraum denkbarer Strafen schrittweise zu verkleinern und dabei zum konkreten Strafmaß zu gelangen. Die Datenbankwerte werden ausdrücklich nur als „allgemeine Hinweise“ bezeichnet, als ein Hilfsmittel, um die übliche Strafzumessungstendenz in vergleichbaren Fällen zu ermitteln. Eine gesetzliche Grundlage für den Einsatz der Datenbank oder gar eine Pflicht zum Einsatz besteht nicht. Dennoch geschieht dies in der Praxis oft; teilweise wird auf den Datenbankwert auch in den Urteilsgründen hingewiesen.

Auch die Staatsanwaltschaft und die Verteidigung haben Zugriff und nutzen die Informationen aus der Datenbank für ihre Anträge bzw. Plädoyers. Schließlich orientie-

60 S. dazu umfassend *Schmidt* 2019.

61 *Schmidt* 2019, 157.

62 Vgl. *Kato* 2019.

ren sich auch die Berufungs- und Revisionsgerichte an der Datenbank. Trotz der fehlenden Verbindlichkeit war eine erhebliche Abweichung vom Datenbankwert in der Vergangenheit bereits Anlass für den OGH, Urteile aufzuheben.⁶³ Hier besteht offensichtlich ein Spannungsverhältnis zwischen dem Anliegen, das bisherige Strafniveau beizubehalten und dem Zweck der Laienbeteiligung, deren (ggf. auch abweichende) Ansichten in die Strafzumessung einfließen zu lassen.

Tatsächlich hat sich das Strafniveau seit Einführung des Saibanⁱⁿ-Systems in bestimmten Bereichen verändert, und zwar in beide Richtungen: die Strafen bei Sexualdelikten sowie bei Totschlag und Körperverletzung mit Todesfolge sind tendenziell schärfer geworden, während sie bei anderen Delikten wie Brandstiftung oder bei speziellen Konstellationen von Mitleids-Tötungen milder geworden sind.⁶⁴ Der mögliche Kritikpunkt einer zu starken „Zementierung“ des bisherigen status quo durch die Datenbank scheint sich also in der Praxis nicht zu bestätigen. Das Niveau entwickelt sich weiter, mit dem Vorteil, dass Abweichungen nun „sehenden Auges“ und mit entsprechender Begründung vorgenommen werden.

Mögliche Datenschutz-Bedenken werden dadurch abgemildert, dass die im System gespeicherten Daten anonymisiert sind. Der Zugang zur Datenbank ist nur an Terminals in Gerichtsgebäuden und nur unter Verwendung eines Passworts möglich. Allerdings bleiben offene Fragen und Kritikpunkte bestehen. So wäre zu überlegen, ob für eine zumindest faktisch so einflussreiche Datenbank, die eine „normative Orientierungsfunktion“⁶⁵ entfaltet, nicht doch eine gesetzliche Grundlage nötig wäre. Dabei könnte dann auch ganz im Sinne der Wesentlichkeitslehre vom Gesetzgeber (und nicht von einer gerichtsinternen Organisation) die ganz zentrale Frage geklärt werden, wer auf welchem Wege über die Auswahl und Gewichtung der Faktoren entscheidet, die in die Datenbank einfließen. Das ist eine wichtige Weichenstellung für die jeweiligen Vergleichsgruppen und damit in gewisser Weise auch präjudiziell für die zukünftige Strafzumessung. Denkbar wäre, zu diesem Zweck eine mit Vertreterinnen und Vertretern der Praxis und der Wissenschaft besetzte unabhängige Strafzumessungskommission einzusetzen, wie es in den USA zur Vorbereitung der Federal Sentencing Guidelines gemacht wurde.⁶⁶ Zwar hat dieser Vorschlag auf dem DJT 2018 ausdrücklich keine Mehrheit gefunden. Die Einsetzung einer Kommission könnte aber für einen Ausgleich der verschiedenen Interessen und einen möglichst breiten, auch gesellschaftlichen Konsens sorgen. Zudem wäre das Verfahren transparenter – gerade daran fehlt es im Moment noch. Zur Datenbank haben nur sehr wenige Verfahrensbeteiligte Zugang, was aus Datenschutzgründen Vorteile mit sich bringt, aber eben dem Anliegen der Transparenz entgegensteht. Es wäre zu diskutieren, warum die interessierte Öffentlichkeit nicht auch in dieser statistisch aufbereiteten Weise erfahren sollte, wie das derzeitige Strafniveau aussieht.

63 Koike 2019; Schmidt 2019, 158.

64 Schmidt 2019, 160 f.

65 S. auch dazu Koike 2019; Schmidt 2019, 158.

66 S. dazu Kaspar 2018, C 77 f.

III. Computergestützte Strafzumessungssoftware

Denkbar ist zudem eine computergestützte Strafzumessungssoftware, die ohne eine Fall-Datenbank auskommt, aber auch noch keine KI benötigt. Ein solches Modell hat bspw. *Giannoulis* vorgeschlagen: Um die Unschärfen der Strafzumessung angemessen abzubilden, nutzt es die Fuzzy-Logik. Die festgelegten Input-Variablen, im genannten Fall „Erfolgsunrecht“, „Handlungsunrecht“ und „Schuld“ werden einer sog. „Fuzzy-fizierung“ unterzogen, d.h. sie werden in einander überlappende Kategorien eingeteilt, um die verschiedenen Schattierungen der Tat angemessen würdigen zu können. Dieselbe Operation wird für die Output-Variablen Freiheits- und Geldstrafe unternommen.⁶⁷ Die Input-Variablen stehen in Interdependenz zueinander.⁶⁸ Dabei ist das Modell insgesamt transparent und nachvollziehbar, sowohl für Anwenderinnen und Anwender als auch für Betroffene.⁶⁹ Der Einsatzbereich des Modells von *Giannoulis* ist bisher beschränkt auf sechs Eigentums- und Vermögensdelikte und auch dort nicht vollkommen universell einsetzbar.⁷⁰ Eine Erweiterung auch auf andere Delikte erscheint aber zumindest denkbar.

IV. Künstliche Intelligenz

Wie sieht es nun mit dem Einsatz von KI im Bereich der Strafzumessung aus? Der Gedanke erscheint zunächst gewöhnungsbedürftig, aber immerhin erleben wir an anderen Stellen bereits Entwicklungen in diese Richtung. So bestehen in Estland bereits konkrete Pläne, bestimmte schematisierbare zivilrechtliche Entscheidungen mit einem Streitwert von bis zu 7000 Euro vollständig mittels KI treffen zu lassen, ganz im Sinne von Legal Tech 3.0.⁷¹ In den USA wurde in einigen Bundesstaaten (mit eher bescheidenem Erfolg)⁷² das von einem privaten Unternehmen entwickelte, selbstlernende Risikoprognose-Tool COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) für die Rückfallprognose eingesetzt, das seit 2017 unter neuem Namen (Equivant)⁷³ operiert. Von dort erscheint der Schritt nicht weit, zumindest in manchen Teilgebieten des Strafrechts, deren Unrechts- und Schuldgehalt einfach quantifizierbar erscheint, auch für die Strafzumessung selbst in ähnlicher Weise vorzugehen, etwa im Bereich einfacher Diebstahlsdelinquenz oder bei Verkehrsstraftaten.⁷⁴ Denkbar wäre dabei ein Decision-Support-System, bei dem die selbstlernenden Algorithmen – in KI-gestützter Fortführung des schon unter D.III. vorgestellten Ansatzes – nur einen Entscheidungsvorschlag berechnen, von dem das Gericht abweichen darf. Allerdings dürf-

67 *Giannoulis* 2014, 372 ff.

68 A. a. O., 381 ff.

69 A. a. O., 393, 400.

70 A. a. O., 420.

71 Der Standard 2019.

72 Siehe dazu bereits oben, Fn. 45.

73 Siehe <https://www.equivalent.com/>.

74 Zu den in der Praxis in diesem Bereich verwendeten Tabellen s. bereits oben Fn. 8.

ten Abweichungen von „wissenschaftlich ermittelten“ bzw. automatisch „berechneten“ Vorentscheidungen schwer fallen bzw. eher selten zu erwarten sein; der Entscheidungsvorschlag hätte eine immense suggestive Wirkung.⁷⁵

Noch weiter ginge dann eine selbstständig durch KI erstellte Strafzumessungsentcheidung, die selbst rechtliche Verbindlichkeit erlangt. Vor allem für den letzteren Fall stellen sich dabei offensichtlich eine ganze Reihe von Fragen: Auf welcher Datenbasis sollen die entscheidenden Algorithmen erstellt werden – sollen die von der Praxis herangezogenen Kriterien und die dadurch erstellten Urteile entscheidend sein? Ist es weiterführend, mit Hilfe von KI die Urteilsbegründungen auszuwerten, um die Entscheidungsregeln herauszufinden und der zukünftigen Strafzumessung zugrunde zu legen? Was ist mit dem bekannten Einwand, dass die geschriebenen Urteilsgründe oft nicht die tatsächlichen sind?⁷⁶

In rechtlicher Sicht wäre u.a. die Frage, ob das Schuldprinzip überhaupt eine Abkoppelung der Schuldwertung von einem Menschen als Entscheider erlaubt. So wie nach h.M. eine nicht-menschliche Entität nicht Adressat eines Schuldvorwurfs sein kann,⁷⁷ könnte man entsprechend argumentieren, dass eine Maschine schon kraft Natur der Sache nicht in der Lage ist, einen Schuldvorwurf zu erstellen bzw. zu erheben. Auch Art. 97 I GG und Art. 101 GG wären zu diskutieren – ist die Garantie der richterlichen Unabhängigkeit bzw. des gesetzlichen Richters mit der Einrichtung eines „Computerrichters“ vereinbar? Zudem dürfte spätestens ab dieser Stufe die Einführung softwaregestützter Strafzumessung auch im Lichte des Art. 103 Abs. 2 GG Probleme aufwerfen, weil die über den bisherigen gesetzlichen Zustand hinausgehende Vereinheitlichung der Strafzumessung zwangsläufig für einzelne Rechtsunterworfenen auch strafscharfend wirkt. Insofern bedürfte sie zumindest einer klaren gesetzlichen Verankerung, einschließlich der Frage der Bindungswirkung und der Abweichungsmöglichkeiten. Auch wäre zu klären, ob eine solche computerisierte Strafzumessung in der Bevölkerung überhaupt Akzeptanz fände⁷⁸ oder ob das Phänomen der „algorithm aversion“⁷⁹ im Sinne einer intuitiven Skepsis gegenüber maschinell erstellten Entscheidungen entgegenstehen würde.

Bei selbstlernenden Systemen erscheint der Vorwurf, hier würde die Entscheidung in einer Art „Black Box“ getroffen,⁸⁰ nachvollziehbar. Allerdings wird vor einer solchen Redeweise auch gewarnt:⁸¹ Damit werde zu sehr von dem Umstand abgelenkt, dass Art und Inhalt der jeweiligen Entscheidungen von den Daten und Verarbeitungs-

75 Eine Parallele zeigt sich, wenn Sachverständige vor Gericht die Gefährlichkeit Angeklagter im Rahmen eines Prognosegutachtens bejahen und dabei auch Punktwerte der entsprechenden Instrumente nennen (z.B. den „PCL-R-Wert“; vgl. *Hare* 2007). S. auch zum „Automation Bias“ *Sommerer* 2020, 71 ff. und generell zum Ankereffekt *Englich* u.a., *Personality and Social Psychology Bulletin* 32 (2006), 188 ff.

76 S. dazu *Rostalski/Völkening* KRiPoZ 2019, 271.

77 S. statt vieler nur *Wessels* u.a. 2019, Rn. 627.

78 S. dazu *Kilian* NJW 2017, 3043, 3050; vgl. a. *Sommerer* 2020, 307 ff.

79 Vgl. nur *Frick* 2015.

80 S. *Herberger* NJW 2018, 2825, 2828.

81 Vgl. dazu näher *Sommerer* 2020, 196 ff.

regeln abhängen, die dem System von Menschen vorgegeben wurden. Das ist auch der Grund, warum der Begriff der „künstlichen Intelligenz“ zum Teil als missverständlich kritisiert wird, eher gehe es um „erweiterte Intelligenz“, die aber auf verantwortliches menschliches Handeln rückführbar sei.⁸² Hinzu kommt, dass das Gericht ja selbst derzeit auch als eine Art „Black Box“ agiert,⁸³ die sich bei ihrer Entscheidung innerhalb eines sehr weit gesteckten Rahmens bewegen kann. Insofern könnte ein einheitliches, regelgebundenes Vorgehen mit Hilfe von KI doch ein Fortschritt sein. Über die vielfältigen technischen, rechtlichen und letztlich auch ethischen Fragen, die all dies aufwirft, wäre allerdings noch gründlich zu diskutieren – auch vor dem Hintergrund, dass Algorithmen bislang nicht vom Staat selbst, sondern überwiegend von privaten Unternehmen entwickelt werden, was zusätzlich die Gefahr mit sich bringt, dass auch die zugrunde liegenden Quellcodes unbekannt und intransparent sind.⁸⁴

E. Fazit und Ausblick

Im digitalen Zeitalter macht der technische Fortschritt auch vor dem Rechtssystem nicht halt. Schon heute sind verschiedene Lösungen etabliert, die sich dem Bereich der Legal Tech zuordnen lassen. Allerdings betrifft der Großteil dieser Entwicklungen bisher außerstrafrechtliche Bereiche. Zudem sind vor allem Lösungen verbreitet, die sich als Legal Tech 1.0 bezeichnen lassen, insbesondere KI wird bisher noch praktisch nicht eingesetzt. Für die Zukunft stellt der Bereich der Strafzumessung einen Sektor dar, der trotz aller Schwierigkeiten auch viel Potential für Legal Tech bietet. Diese lässt sich insbesondere auch nutzbringend einsetzen, um zu einer wünschenswerten Vereinheitlichung der Strafzumessungspraxis zu gelangen. Hierzu bieten sich zunächst insbesondere Strafzumessungsdatenbanken nach japanischem Vorbild an, als weitere Entwicklungsschritte wäre auch an die Integration von „deliberative polls“ durch Online-Panels zu denken und in letzter Konsequenz an KI-Lösungen zur Vorbereitung oder gar automatisierten Umsetzung der Strafzumessungsentscheidung. Freilich stellen sich insbesondere im letzteren Fall schwierige normative Fragen, und auch die gesellschaftliche Akzeptanz ist ungeklärt. Zudem ist auch denkbar, über eine genauere Untersuchung der so zusammengestellten Faktoren und deren Bedeutung für präventive Zwecke zu passgenaueren, in ihrer Findung nachvollziehbaren Entscheidungen für den

82 S. Herberger NJW 2018, 2825, 2827.

83 S. jedoch kritisch zu einer Vergleichbarkeit von menschlicher und algorithmischer Intransparenz, Sommerer 2020, 226 ff.

84 Davor eindringlich warnend Sommerer 2020, 196 ff.

Einzelfall zu kommen. Weitere empirische und normative Forschung in diesem Bereich ist allerdings dringend erforderlich.

Literatur

Angwin/Larsen/Mattu/Kirchner (2016) Machine Bias – There’s software used across the country to predict future criminals. And it’s biased against blacks, ProPublica (<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>)

Cochran/Chamlin Can Information Change Public Opinion? Another Test of the Marshall Hypotheses, in: *Journal of Criminal Justice* 33 (2005), 573-584

Coester (2018) Das Düsseldorfer Gutachten und die Folgen, in: Walsh u.a. (Hrsg.), *Evidenzorientierte Kriminalprävention in Deutschland*, 37

Danziger/Levav/Avnaim-Pesso Extraneous factors in judicial decisions, in: *PNAS* 2011, 6889-6892 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3084045/>)

Der Standard (2019) Euer Ehren Roboter. Estland will Richter durch künstliche Intelligenz ersetzen (www.derstandard.at/story/2000100613536/justiz-estland-will-richter-durch-kuenstliche-intelligenz-ersetzen)

Dölling Generalprävention durch Strafrecht: Realität oder Illusion?, in: *ZStW* 102 (1990), 1-20

Dreher Anmerkung zu OLG Stuttgart, Urteil vom 26.8.1960, in: *MDR* 1961, 343-344

Dressel/Farid The Accuracy, Fairness, and Limits of Predicting Recidivism, in: *Science Advances* 4 (2018)

Eierle (2019) „Verteidigung der Rechtsordnung“ und „öffentliches Interesse“ als Rechtsbegriffe mit empirischem Gehalt?, in: Kaspar/Walter (Hrsg.), *Strafen „im Namen des Volkes“?*, 163

Englich/Mussweiler/Strack Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts’ Judicial Decision Making, in: *Personality and Social Psychology Bulletin* 32 (2006), 188-200

Frick Here’s Why People Trust Human Judgement Over Algorithms, in: *Harvard Business Review* 2015 (<https://hbr.org/2015/02/heres-why-people-trust-human-judgment-over-algorithms>)

Giannoulis (2014) *Studien zur Strafzumessung*

Green (2008) *When Children Kill Children*

Grob (2019a) *Künstliche Intelligenz*, in: Weber (Hrsg.), *Creifelds, Rechtswörterbuch*

- Grob* (2019b) Legal Tech, in: Weber (Hrsg.), Creifelds, Rechtswörterbuch
- Hamann* (2014) Evidenzbasierte Jurisprudenz
- Hare* (2007), Hare Psychopathy Checklist-Revised (PCL-R): Technical Manual, 2. Aufl.
- Harrendorf* Sentencing Thresholds in German Criminal Law and Practice: Legal and Empirical Aspects, in: Criminal Law Forum 28 (2017), 501-539
- Harrendorf* (2018) Zur Adäquität von Strafe bei der Beförderungerschleichung, in: Dünkel u.a. (Hrsg.), Gedächtnisschrift für Joecks, 97
- Harrendorf* (2020) Absolute und relative Bagatellen. Grenzen des Strafrechts bei geringfügiger Delinquenz (im Erscheinen)
- Harrendorf/Geng* (2019) Der rational kalkulierende Verbrecher? Zur Entwicklung, Stand und Zukunftsperspektiven ökonomischer Kriminalitätstheorien, in: Körnert/Lege/Grube (Hrsg.), Recht trifft Wirtschaft, 189
- Heinz* Valide und aussagekräftige statistische Erfassung von Kriminalität und Kriminalitätskontrolle – eine notwendige Bedingung für evidenzbasierte Kriminalpolitik, in: NK 2020, 3.
- Herberger* „Künstliche Intelligenz“ und Recht, in: NJW 2018, 2825-2828
- Hieda* Strafzumessungs-Datenbank und Strafzumessung, Vortrag auf dem Deutsch-Japanischen Symposium in Tokio am 22.9.2019 (bislang unveröffentlicht)
- Hirsch/Jareborg* (1991) Strafmaß und Strafgerechtigkeit
- Höffler* (2018) „Evidence based“ Kriminalpolitik?, in: Zabel (Hrsg.), Strafrechtspolitik, 225
- Höffler* (2019) Die Herausforderungen der globalisierten Kriminalität an die Kriminologie – am Beispiel Risikoprognosen, in: Dessecker/Harrendorf/Höffler (Hrsg.) Angewandte Kriminologie – Justizbezogene Forschung, 51
- Hörnle* (1999) Tatproportionale Strafzumessung
- Indermaur/Roberts/Spiranovic/Mackenzie/Gelb* A Matter of Judgement: The Effect of Information and Deliberation on Public Attitudes to Punishment, in: Punishment & Society 14 (2012), 147-165
- Kaspar* (2014) Verhältnismäßigkeit und Grundrechtsschutz im Präventionsstrafrecht
- Kaspar* (2018) Sentencing Guidelines vs. Freies richterliches Ermessen. Gutachten C für den 72. Deutschen Juristentag
- Kaspar* (2019) Wege zur Strafgleichheit, in: Beckmann u.a. (Hrsg.), Gedächtnisschrift für Tröndle, 279
- Kaspar* Herausforderungen evidenzbasierter Strafgesetzgebung, in: KriPoZ 2020, 14-21

Kaspar/Walter (Hrsg.) (2019) Strafen „im Namen des Volkes“?

Kato Strafzumessungs-Datenbank und Strafzumessung aus japanischer strafverfahrensrechtlicher Sicht, Vortrag auf dem Deutsch-Japanischen Symposium in Tokio am 22.9.2019 (bislang unveröffentlicht)

Kilian Die Zukunft der Juristen, in: NJW 2017, 3043-3050

Kleinberg/Lakkaraju/Leskovec/Ludwig/Mullainathan Human Decisions and Machine Predictions, in: The Quarterly Journal of Economics 133 (2018), 237-293

Koike Strafzumessungsdatenbank und Strafzumessungsrecht aus Sicht des materiellen Rechts in Japan, Vortrag auf dem Deutsch-Japanischen Symposium in Tokio am 22.9.2019 (bislang unveröffentlicht)

V. Liszt (1905) Kriminalpolitische Aufgaben, in: Strafrechtliche Aufsätze und Vorträge, Band I, 290

Luskin/Fishkin/Jowell Considered Opinions: Deliberative Polling in Britain, in: British Journal of Political Science 32 (2002), 455-487

Martini (2019) Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz

Maurer (2005) Komparative Strafzumessung

Miebach/Maier (2016) § 46 StGB, in: Joecks/Miebach (Hrsg.), Münchener Kommentar zum StGB, 3. Aufl.

Müller-Dietz (1985) Integrationsprävention und Strafrecht. Zum positiven Aspekt der Generalprävention, in: Vogler u.a. (Hrsg.), Festschrift für Jescheck, 813

Roberts/Harrendorf (2020) Criminal History Enhancements at Sentencing, in: Ambos u.a. (Hrsg.), Core Concepts in Criminal Law and Criminal Justice, Vol. 1, 261

Robinson (2014) Intuitions of Justice and the Utility of Desert

Rössner/Bannenberg/Coester/Erwin/Leonhardt (2002) Düsseldorfer Gutachten: Empirisch gesicherte Erkenntnisse über kriminalpräventive Wirkungen

Rostalski/Völkening Smart Sentencing – Ein neuer Ansatz für Transparenz richterlicher Strafzumessungsentscheidungen, in: KriPoZ 2019, 265-273

Russel/Norvig (2016) Artificial Intelligence: A Modern Approach

Schäfer/Sander/van Gemmeren (2017) Praxis der Strafzumessung, 6. Aufl.

Schmidt (2019) Das japanische Saiban‘in System und das deutsche Schöffensystem

Schöch (1973) Strafzumessungspraxis und Verkehrsdelinquenz

Schulte (2016) Die Methode der richterlichen Straftatenprävention

Sherman/Farrington/Welsh/MacKenzie (2002) Evidence-Based Crime Prevention

Singelstein Predictive Policing: Algorithmenbasierte Straftatenprognosen zur vorausschauenden Kriminalintervention, in: NSStZ 2018, 1-8

Sommerer (2020) Personenbezogenes Predictive Policing. Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose (im Erscheinen)

Spirgath (2013) Zur Abschreckungswirkung des Strafrechts

Streng (1984) Strafzumessung und relative Gerechtigkeit

Walsh/Pniewski/Kober/Armborst (Hrsg.) (2018), Evidenzorientierte Kriminalprävention in Deutschland

Weigend (1988) Richtlinien für die Strafzumessung, in: Rüfer (Hrsg.), Festschrift der Rechtswissenschaftlichen Fakultät zur 600-Jahr-Feier der Universität zu Köln, 579

Wessels/Beulke/Satzger (2019) Strafrecht Allgemeiner Teil, 49. Aufl.

Kontakt

Prof. Dr. Johannes Kaspar

Lehrstuhl für Strafrecht, Strafprozessrecht, Kriminologie und Sanktionenrecht

Universität Augsburg

johannes.kaspar@jura.uni-augsburg.de

Prof. Dr. Katrin Höffler

Abteilung für Kriminologie, Jugendstrafrecht und Strafvollzug

Lehrstuhl für Strafrecht und Kriminologie

Universität Göttingen

katrin.hoeffler@jura.uni-goettingen.de

Prof. Dr. Stefan Harrendorf

Lehrstuhl für Kriminologie, Strafrecht, Strafprozessrecht und vergleichende Strafrechtswissenschaften

Universität Greifswald

stefan.harrendorf@uni-greifswald.de