

Smart Sentencing

Ein neuer Ansatz für Transparenz richterlicher Strafzumessungsentscheidungen

von Prof. Dr. Dr. Frauke Rostalski
und Malte Völkening*

Abstract

Vergleichende Untersuchungen zur Strafzumessungspraxis in Deutschland ergeben regelmäßig das Bild starker Variationen in der Strafhöhe, die sich nicht ausschließlich durch normativ relevante Unterschiede der betroffenen Fälle erklären lassen. Wird die Strafzumessung in der Gesellschaft und durch den Verurteilten als willkürlich empfunden, so schwindet die Akzeptanzerwartung der Entscheidungen. Abstrakte Strafzumessungsrichtlinien nach U.S.-amerikanischem Vorbild versprechen Abhilfe, sind aber nicht hinreichend auf die Umstände des Einzelfalls konkretisierbar. Vorzugswürdig erscheint, das als angemessen empfundene Strafmaß durch einen kritischen und detailgenauen Vergleich mit ähnlichen Entscheidungen zu ermitteln. Dazu ist aber eine umfassende und transparente Datengrundlage erforderlich. Eine solche kann mittels Legal Tech automatisiert anhand der Zumessungserwägungen in den Begründungen aller einschlägigen Urteile erstellt werden.

Comparative studies regularly reveal strong variance in the level of penalties imposed by German judges that is not fully explicable by actual differences between the cases in question. If verdicts are seen as arbitrary by the society or the defendants, their prospect of acceptance decreases. Sentencing guidelines, as used in the U.S., appear to remedy these deficits but they lack individualization. Therefore, a critical and detailed comparison with similar cases and decisions seems more promising. This approach requires an extensive and transparent empirical basis which can be provided by analyzing the reasons given by other courts, a process that can be automated using Legal Tech.

I. Einleitung

„Der Richter pflegt [...] bei der Strafzumessung bewußt oder unbewußt nicht vom gesetzlichen Strafrahmen, sondern von einem gewissen typischen Testfall auszugehen,

für den er eine bestimmte Strafe für angemessen hielt. Und an diesem Fall und an dieser Strafe mißt er dann die kommenden Fälle und paßt die Strafe für sie jener Strafe an, die als Fixpunkt seiner gesamten Strafzumessung erscheint. Sie ist sein geheimes Metermaß.“¹ Diese Worte von Eduard Dreher haben heute nicht viel von ihrer Gültigkeit verloren.² Über die Legitimität eines solchen Verfahrens ist damit freilich noch nichts gesagt.³ Hier stellt sich einerseits die Frage, ob die Gerichte durch das Anlegen eigener Wertmaßstäbe in unzulässiger Weise von den gesetzlichen Vorgaben abweichen und sich so an die Stelle des Gesetzgebers setzen.⁴ Andererseits hat schon Dreher erkannt, dass das „geheime Metermaß“ der Richter ein lokal beschränktes ist, dass sich also unterschiedliche regionale Tendenzen in den Strafzumessungsentscheidungen herausbilden.⁵ Dieses Problem wird spätestens seit dem 72. Deutschen Juristentag 2018 wieder intensiv diskutiert.⁶ Dabei sind die Lösungsstrategien seit Jahren im Wesentlichen unverändert geblieben. Insbesondere der Ruf nach „Sentencing Guidelines“ entsprechend U.S.-amerikanischem Vorbild wird vielerorts erhoben.⁷ Bislang bei weitem nicht ausgeschöpft erscheint den Autoren dieses Beitrags dabei das Potenzial von Legal Tech für die Vereinheitlichung der richterlichen Strafzumessungsentscheidungen.

II. Strafungerechtigkeit als Gefahr für die gesellschaftliche Anerkennung von Strafurteilen

Die Strafzumessungsentscheidungen in Deutschland unterscheiden sich mitunter erheblich. Angesichts der stets einzubeziehenden Besonderheiten des Einzelfalls verwundert das im Grunde wenig. Als problematisch erweisen sich Strafzumessungsunterschiede allerdings dann, wenn sie auf keinem normativ relevanten Umstand beruhen. Dies ist u.a. dann der Fall, wenn sich die Abweichungen im Wesentlichen durch die individuelle Person des entscheidenden Richters oder aber den Ort erklären, an

* Prof. Dr. Dr. Frauke Rostalski ist Inhaberin des Lehrstuhls für Strafrecht, Strafprozessrecht, Rechtsphilosophie und Rechtsvergleichung an der Universität zu Köln. Malte Völkening ist Wissenschaftlicher Mitarbeiter an diesem Lehrstuhl.

¹ Dreher, MDR 1961, 343 (344).

² Siehe unten unter II.

³ Dazu unten unter IV.

⁴ So der Vorwurf bei Dreher, MDR 1961, 343 (344).

⁵ Dreher, MDR 1961, 343 (344).

⁶ Siehe instruktiv Kaspar, Sentencing Guidelines versus freies tatrichterliches Ermessen – Brauchen wir ein neues Strafzumessungsrecht?, 2018.

⁷ Hoven, KriPoZ 2018, 276 (289 f.); Weigend, in: FS Rechtswissenschaftliche Fakultät Köln, 1988, S. 579 (601).

dem die Entscheidung getroffen wurde.⁸ Weder die Rechtsanwenderpersönlichkeit noch der Entscheidungsort dürfen für die Strafzumessung relevant sein. Vielmehr gebietet es der Grundsatz der rechtlichen Gleichbehandlung sachlich gleich gelagerter Fälle, diese für die Strafhöhenbemessung sachfremden Kriterien außer Acht zu lassen.⁹

Dennoch spielen gerade diese Faktoren den Angaben der Verfasser unterschiedlicher Studien zufolge im Bereich der deutschen Strafzumessung nach wie vor eine alles andere als untergeordnete Rolle.¹⁰ So besteht bisweilen eine erhebliche Diskrepanz der Strafzumessung zwischen verschiedenen Regionen und Gerichtsbezirken.¹¹ Unterschiede zeigen sich vor allem entlang der Nord-Süd-Achse.¹² Innerhalb der einzelnen Gerichtsbezirke erfolgt die Strafzumessung dagegen vergleichsweise homogen.¹³ Mit dem Befund des regionalen Abweichens von Strafzumessungsentscheidungen gehen Schwierigkeiten einher, die in letzter Konsequenz die rechtsstaatlich relevante Akzeptanz von Strafurteilen gefährden. Die Bestrafung dient der (Wieder-)Herstellung des Rechtsfriedens. Durch die Straftat stellt der Täter das Recht infrage. Im Wege des Schuldspruch und ggf. der zusätzlichen Verhängung eines Strafübels reagiert die rechtlich verfasste Gesellschaft hierauf und teilt dem Einzelnen auf diese Weise mit, dass das Recht trotz seiner Tat weiterhin uneingeschränkte allgemeine Gültigkeit hat.¹⁴ Dabei orientieren sich Schuldspruch und Strafübel ausschließlich an dem individuellen, hinreichend gewichtigen Verhaltensnormverstoß des Täters (ggf. nebst Fehlverhaltensfolgen).¹⁵ In diesem Sinne hat die Bestrafung *tatproportional* zu erfolgen:¹⁶ Ihr Gegenstand ist die unberechtigte Freiheitsanmaßung des Täters in Gestalt seiner Straftat, die durch Strafe ausgeglichen wird. Vor diesem Hintergrund lässt sich weder eine zu niedrige noch eine zu hohe Strafe rechtfertigen. Insbesondere dürfen Faktoren, die keinen rechtlich relevanten Bezug zu der Straftat aufweisen, bei der Strafzumessung

nicht berücksichtigt werden.¹⁷ In der Folge spielen das (willkürliche) Belieben der individuellen Richterpersönlichkeit ebenso wie der Entscheidungsort für die Strafzumessung grundsätzlich keine Rolle. Beide treffen für sich genommen keine Aussage über das Gewicht des konkreten Verhaltensnormverstoßes (ggf. nebst Fehlverhaltensfolgen). Sofern diese Faktoren gleichwohl einen (ggf. erheblichen) Einfluss auf die konkrete Strafzumessungsentscheidung haben, beeinträchtigt dies die Akzeptanz des Urteils sowohl durch den Verurteilten selbst als auch durch Dritte. Die Einbeziehung sachfremder Erwägungen erweist sich als willkürlich und bedroht daher das Rechtssicherheitsgefühl. Die (Wieder-)Herstellung des Rechtsfriedens durch Schuldspruch und Strafe kann allein gelingen, wenn der jeweilige Rechtsfolgenausspruch als angemessene Reaktion auf die Tat wahrgenommen wird. Ist dies nicht der Fall, droht er, einen gänzlich gegenteiligen Effekt hervorzurufen: Eine als willkürlich empfundene Entscheidung lässt das Vertrauen in das Funktionieren der staatlichen Ordnung schwinden. Vor allem, wenn dies eine so eingriffsintensive Maßnahme wie staatliche Strafe betrifft, birgt dies eine besondere Gefahr für den gesellschaftlichen Zusammenhalt.¹⁸

III. Fehlen konkreter gesetzlicher Vorgaben für die Strafzumessung im Einzelfall und Vorschlag der Einführung von Strafzumessungsrichtlinien

Wer nach Gründen für diesen Missstand im Bereich der deutschen Strafzumessungspraxis sucht, mag geneigt sein, auf die Weite der Strafrahmen der Delikte im Besonderen Teil des StGB oder aber auf die Regelung des § 46 Abs. 1 S. 1 StGB zu verweisen. Weder das eine noch das andere enthält konkrete Angaben für die angemessene Strafhöhe spezifischer Einzelfälle.¹⁹ Trotz mitunter berechtigter Kritik, sowohl an den Deliktsstrafrahmen²⁰ als auch der strafzumessungsrechtlichen Kernvorschrift²¹,

⁸ Gemeint sind hierbei freilich sachlich vergleichbare Fälle. Beispielsweise die besondere Brutalität der Tat spielt eine Rolle für deren Bewertung. Sollte diese besonders häufig in einer bestimmten Region Deutschlands auftreten, stellt dies selbstverständlich einen normativ unproblematischen Grund für die häufig strengere Bestrafung dar. Siehe *Meier*, Strafrechtliche Sanktionen, 4. Aufl. (2015), S. 255 sowie *Albrecht*, ZStW 102 (1990), 596 (598 f.) dazu, dass sich ein nicht unerheblicher Teil regionaler Unterschiede auf solche normativ relevanten Faktoren zurückführen lässt.

⁹ Vgl. zum Verstoß gegen Art. 3 Abs. 1 GG durch regional abweichende Praxen *Albrecht*, Strafzumessung bei schwerer Kriminalität, 1994, S. 21 ff.; *Kaspar*, Sentencing Guidelines (Fn. 6), C 86 ff. *Streng*, Strafrechtliche Sanktionen, 3. Aufl. (2012), Rn. 504 sieht einen Verstoß „zumindest gegen die *Idee* des verfassungsrechtlichen Gleichbehandlungsgebots“ (Herv. im Original).

¹⁰ Siehe nur die Nachweise bei *Kaspar*, Sentencing Guidelines (Fn. 6), C 18 ff.

¹¹ *Meier*, Sanktionen (Fn. 8), S. 257 m.w.N. Skeptisch für die von ihm untersuchten schweren Delikte aber *Albrecht*, ZStW 102 (1990), 596 (614).

¹² Allerdings mit Ausnahme Baden-Württembergs, vgl. *Grundies*, in: Hermann/Pöge, Kriminalsoziologie, 2018, S. 295 (302 f.).

¹³ Für die Trunkenheit im Verkehr *Schöch*, Strafzumessungspraxis und Verkehrsdelinquenz, 1973, S. 125 ff. Auch *Grundies*, in: Hermann/Pöge, S. 295 (303) sieht einen Zusammenhang zwischen Strafhöhenvarianz und OLG-Bezirkszugehörigkeit, den er mit der „hierarchische[n] Struktur der Justiz“ und der „Revisionspraxis“ erklärt. Angesichts der nur begrenzten Zuständigkeit der Oberlandesgerichte als Revisionsgerichte (§ 121 Abs. 1 Nr. 1 GVG) ist dies aber nur für Entscheidungen in Fällen leichter und mittlerer Kriminalität plausibel (vgl. § 24 Abs. 1 S. 1 Nr. 2 GVG).

¹⁴ *Rostalski*, Der Tatbegriff im Strafrecht, 2019, S. 25 f.

¹⁵ *Rostalski*, Tatbegriff (Fn. 14), S. 27, 133 f., 139; *Freund*, GA 1999, 509 (526 f.).

¹⁶ Siehe zum Konzept statt aller *Hörnle*, Tatproportionale Strafzumessung, 1999.

¹⁷ Siehe allerdings bereits *Timm*, Gesinnung und Straftat, 2012, S. 256 ff. dazu, dass die Vorschrift des § 46 StGB auch gegenwärtig Strafzumessungsfaktoren auflistet, die sich normativ nicht rechtfertigen lassen.

¹⁸ Vgl. zur hohen Sensibilität des Gleichheitsempfindens bei Strafzumessungsentscheidungen *Streng*, Sanktionen (Fn. 9), Rn. 505 m.w.N.

¹⁹ *Freund*, GA 1999, 509 (524), der im Hinblick auf die Strafrahmen auf die (wenigen) Ausnahmen einer absoluten Strafdrohung und der Anordnung einer Mindeststrafe für bestimmte Fälle verweist.

²⁰ *Kinzig*, in: Schönke/Schröder, 30. Aufl. (2019), § 46 Rn. 72 plädiert für eine „deutlichere Konturierung der Strafrahmen“, *Streng*, in: Kindhäuser/Neumann/Paeffgen, 5. Aufl. (2017), § 46 Rn. 200 für eine Absenkung der Strafrahmenobergrenzen, wobei letzterer zugeibt, dass hiervon nur eine geringfügige Verbesserung zu erwarten ist.

²¹ Kritisch zum unklaren Verhältnis der in § 46 Abs. 2 StGB genannten Kriterien etwa *Kaspar*, Sentencing Guidelines (Fn. 6), C 12, 66 ff; *Kudlich/Koch*, NJW 2018, 2762 (2763).

ginge allerdings die Forderung deutlich zu weit, diese hätten konkrete Lösungen für den jeweiligen Fall zu liefern, über den der Strafrichter entscheiden muss. So ist es gerade als ein Vorzug des deutschen Systems zu bewerten, dass weite Strafraumen ein breites Reaktionsinstrumentarium an die Hand geben.²² Ebenso verhält es sich im Hinblick auf die Vorschrift des § 46 Abs. 1 S. 1 StGB, soweit diese die Strafzumessung an der „Schuld“ des Täters orientiert wissen will. Die Norm lässt sich ohne Weiteres so verstehen, dass durch die Bestrafung ein Ausgleich für die begangene Straftat in Gestalt des individuellen, hinreichend gewichtigen Verhaltensnormverstoßes (ggf. nebst Fehlverhaltensfolgen) geschaffen werden soll. Sie entspricht damit in dieser Hinsicht unmittelbar den bereits dargelegten straftheoretischen Vorgaben.

Und dennoch richten nicht wenige ihre Verbesserungsvorschläge zur Beseitigung der unliebsamen gegenwärtigen Strafzumessungspraxis auf das Gesetz bzw. zumindest dessen Konkretisierung. So erleben wir derzeit in der strafrechtswissenschaftlichen Debatte eine (neuerliche) Konjunktur der Forderung nach einer mehr oder minder stark ausgeprägten Orientierung am Vorbild der U.S.-amerikanischen Strafzumessungsrichtlinien („Sentencing Guidelines“).²³ Diese sehen eine gegenüber der Vorschrift des § 46 StGB erhöhte Ausdifferenzierung relevanter Strafzumessungsfaktoren in Tabellenform vor, die der Richter bei der Begründung seiner Strafzumessungsentscheidung heranziehen kann. In den USA werden Strafzumessungsrichtlinien von einer unabhängigen Expertenkommission erarbeitet und regelmäßig angepasst. Sie enthalten detaillierte Vorgaben, anhand deren sich aus dem Strafraumen eines Delikts unter Berücksichtigung strafschärfender und strafmildernder Umstände ein vergleichsweise präzises Strafmaß ergibt.²⁴ Der Fokus liegt dabei auf den Tatumständen und der Vorstrafenbelastung des Täters.²⁵ Während die Befolgung der Richtlinien anfangs zwingend war, sind sie seit einem Urteil des Supreme Court aus dem Jahr 2005 nur noch unverbindliche Anhaltspunkte, mit denen sich die Richter allerdings im Urteil auseinandersetzen müssen.²⁶ In Deutschland wird vor diesem Hintergrund seitens der Befürworter der Einführung von „Sentencing Guidelines“ u.a. darüber diskutiert, inwieweit eine Orientierung des Richters an den Strafzumessungsrichtlinien verbindlich sein soll, ob bzw. mit welchem zusätzlichen Dokumentationsaufwand Abweichungen zulässig sein sollen und welche Vor- oder Nachteile mit einem Punktesystem einhergingen. Einig ist man sich zumindest im Hinblick auf die Vorzüge eines solchen Systems gegenüber der gegenwärtigen Strafzumessungspraxis: Mehr Transparenz soll die Gleichheit von Straf-

maßentscheidungen befördern, was zuletzt der Strafgerechtigkeit diene.²⁷

Der Idee einer Einführung von Strafzumessungsrichtlinien werden seit Langem bedeutsame Einwände entgegengehalten. So erscheint bereits fraglich, ob der große Aufwand der Erstellung solcher Richtlinien noch in einem angemessenen Verhältnis zu ihrem Nutzen steht.²⁸ Als problematisch erweist sich zumindest das schematische Vorgehen, das entsprechende Richtlinien dem Rechtsanwender im Rahmen der Strafzumessungsentscheidung auferlegen. Will er sich hieran *nicht* halten, bedarf es jedenfalls im U.S.-amerikanischen Modell einer besonderen Begründung, was eine neuerliche Hürde darstellen kann in der Würdigung des Einzelfalls (sofern der Richter die erhöhte Mühe scheut).²⁹ Strafzumessungsrichtlinien, die einem solchen schematischen Denken verhaftet sind, machen die Regel – die Besonderheit des Einzelfalls – zum Ausnahmefall. Hiermit geht die Gefahr einher, dass die individuelle Gerechtigkeit auf der Strecke bleibt.³⁰ Hinzu kommt, dass sich die abstrakte Umwandlung von Zumesungserwägungen in eine bestimmte Strafe besonders für solche Merkmale eignet, die häufig vorkommen und quantifizierbar sind (etwa die Schadenshöhe beim Diebstahl oder die BAK bei der Trunkenheit im Verkehr). Ein System abstrakter Richtlinien läuft daher Gefahr, sich auf derartige Merkmale zu konzentrieren und weniger gut schematisierbare Kriterien zu vernachlässigen, wengleich diese im Einzelfall eine wesentlich höhere Bedeutung haben können.³¹

IV. Vergleichende Strafzumessung als Schlüssel zu mehr Strafgerechtigkeit

Wer so auf das Problem bestehender Missstände im Bereich der deutschen Strafzumessungspraxis blickt, kann den Eindruck gewinnen, die Wahl zwischen zwei Übeln treffen zu müssen: Dem Festhalten am bisherigen, die Praxis offenbar unzureichend anleitenden Modell auf der einen, der Einführung von irgendwie gearteten Strafzumessungsrichtlinien auf der anderen Seite. Indessen zeichnet die bislang dargelegte Kritik lediglich ein verkürztes Bild vom Spielraum desjenigen, der sich auf die Suche nach mehr Strafzumessungsgerechtigkeit begibt. So liegt der Schlüssel zur Lösung des Problems in einem Ausschöpfen der durch das Gesetz bereits in seiner gegenwärtigen Fassung eingeräumten Möglichkeiten. Diese können unter Umständen selbst durch geeignete Strafzumessungsrichtlinien konkretisiert werden, soweit diese keinen sachlich unangemessenen Schematismus in den Vorgang der Strafzumessung hineinragen – was nicht zwingend ist.³²

²² Siehe dazu bereits *Freund*, GA 1999, 509 (515), der darauf verweist, dass ein vom gegenwärtigen System der Strafraumen abweichendes „einseitiges Rechtssicherheitsdenken“ die Einzelfallgerechtigkeit in unerträglicher Weise vernachlässigen würde.

²³ Siehe zu früheren Vorschlägen beispielsweise *Weigend*, in: FS Rechtswissenschaftliche Fakultät Köln, 1988, S. 579 (591 ff.) und *Reichert*, *Intersubjektivität durch Strafzumessungsrichtlinien*, 1999, S. 273 ff. sowie aus der jüngeren Zeit *Hoven*, KriPoZ 2018, 276 (289 f.); affirmativ auch *Wohlers/Went*, in: *Hirsch/Neumann/Seelmann*, *Strafe – Warum?*, 2011, S. 173 (202 f.).

²⁴ Ausführlich dazu *Reichert*, *Strafzumessungsrichtlinien* (Fn. 23), S. 204 ff.

²⁵ *Reichert*, *Strafzumessungsrichtlinien* (Fn. 23), S. 205.

²⁶ *United States v. Booker*, 125 S. Ct. 738 (2005). Siehe dazu sowie zur Kritik an dieser Entscheidung *Meyer*, ZStW 118 (2006), 512 (534 ff.).

²⁷ Vgl. stellvertretend *Hoven*, KriPoZ 2018, 276 (289 f.).

²⁸ *Kaspar*, *Sentencing Guidelines* (Fn. 6), C 84; *Kudlich/Koch*, NJW 2018, 2762 (2765).

²⁹ Zur Hürde, die selbst eine bloße Orientierungshilfe für eine Abweichung darstellt, auch *Kudlich/Koch*, NJW 2018, 2762 (2765 Fn. 25).

³⁰ *Streng*, in: *Kindhäuser/Neumann/Paeffgen*, § 46 Rn. 199.

³¹ Vgl. *Kudlich/Koch*, NJW 2018, 2762 (2764 f.).

³² Vgl. *Hoven*, KriPoZ 2018, 276 (290).

In erster Linie muss aber die bei der Strafhöhenbemessung unbedingt anzuwendende *Technik des Vergleichs mit anderen Fällen* der zu beurteilenden Straftatbegehung verbessert werden. Gerechte Strafzumessung kann allein unter der Voraussetzung gelingen, dass der Einzelfall in ein Verhältnis zu früheren Entscheidungen vergleichbarer Fälle gebracht wird, die sich mit den rechtlich anerkenntenswerten gesellschaftlichen Wertvorstellungen decken und daher Akzeptanz hervorrufen.³³ Wer von einem vergleichbaren Fall der Deliktsbegehung als „Fixpunkt“ ausgeht, von dem aus er Abweichungen nach oben oder unten vornehmen kann, besitzt einen festen Ausgangspunkt, an dem er seine Entscheidung orientieren kann.³⁴ Dabei erweist es sich als praktisch sinnvoll, sich jedenfalls nicht an nur selten vorkommenden Fällen auszurichten, sondern solche zu wählen, die mit entsprechenden strafzumessungsrelevanten Eigenschaften schon oft entschieden worden sind. Nimmt der Richter in seinem konkreten Fall hiervon Abweichungen vor, muss er diese begründen und dabei an den Umständen festmachen, die die von ihm zu bewertende Tat kennzeichnen. Allein auf diese Weise kann ein angemessener Einstieg in den Vorgang der Strafzumessung sowie eine Berücksichtigung der Spezifika des Einzelfalls gelingen.³⁵

Auch der Rechtsprechung lässt sich die Relevanz des Vergleichs mit früheren Entscheidungen innerhalb der Strafhöhenbemessung im Einzelfall entnehmen. Danach liegt ein Rechtsfehler vor, wenn ein Instanzengericht eine verglichen mit dem „üblichen [sic]“ „außergewöhnlich hohe Strafe“ verhängt, ohne dies gesondert zu rechtfertigen.³⁶ Eine Orientierung an Erwägungen anderer Richter soll insbesondere im Bereich der Massenkriminalität zulässig sein.³⁷ Gleichwohl werden gegen das Modell der komparativen Strafzumessung Einwände erhoben. Diese klingen bereits unterschwellig in der Forderung an, die Fälle müssten stets vergleichbar sein.³⁸ Auch betont der *BGH*, dass der Strafrichter durch den Vergleich nicht von der Pflicht befreit wird, sich eine eigene Meinung über die Schuld des Angeklagten und die angemessene Strafe zu bilden.³⁹ Zudem konstatiert der *BGH*, dass eine umfassende Würdigung im Sinne von § 46 Abs. 1 StGB die Berücksichtigung sämtlicher Nuancen des Einzelfalls verlangt.⁴⁰ Dies sei durch den bloßen Vergleich des eigenen Falls mit der Urteilsbegründung in einem anderen Fall aber kaum möglich. Denn einerseits gibt letztere nicht alle Strafzumessungserwägungen wieder, sondern nur solche,

die schlussendlich „bestimmend“ waren (§ 267 Abs. 3 S. 1 Hs. 2 StPO).⁴¹ Andererseits ist aus der Begründung nicht ohne Weiteres ersichtlich, wie stark einzelne Kriterien ausgeprägt waren, beispielsweise wie glaubhaft und umfassend ein Geständnis des Täters tatsächlich war.⁴²

Dem *BGH* ist im Grundsatz zuzustimmen. Allerdings handelt es sich bei dessen Ausführungen bei näherem Hinsehen um keine echte Kritik am Konzept der komparativen – vergleichenden – Strafzumessung. Es ist eine Selbstverständlichkeit, dass Unvergleichbares nicht verglichen werden kann – aus dem dabei allein möglichen Abgleich können keine Rückschlüsse für den jeweils anderen Fall gezogen werden (außer eben demjenigen, dass eine Orientierung am Abgleichsobjekt nicht möglich ist). Der *BGH* bringt insofern treffend zum Ausdruck, worauf es bei einem sachlich angemessenen Vergleich ankommt: Die *Vergleichbarkeit* der Fälle. Es ist aber mitnichten davon auszugehen, dass diese unter keinen Umständen je besteht. Sie ist vielmehr durchaus dann gegeben, wenn sich der Vergleich auf diejenigen Aspekte beschränkt, in denen die Sachverhalte wertend betrachtet übereinstimmen. Geht es beispielsweise um zwei Fälle eines einfachen Diebstahls, bei denen jeweils ein Schaden in Höhe von 50 € verursacht wurde, wohingegen der Täter in einem Fall planvoll vorging, im anderen Fall nur die bloße Gunst eines unbeobachteten Augenblicks ausnutzte, so beschränkt sich die Vergleichbarkeit auf die Schadenshöhe. Infolge der Unterschiede bei der festzustellenden kriminellen Energie kann sich das Strafmaß des zweiten nicht einfach am Strafmaß des ersten Falls orientieren. Dennoch hat letzterer eine (beschränkte) Aussagekraft: Im Vergleich mit weiteren Fällen lässt sich ihm entnehmen, welchen Einfluss eine Schadenshöhe von 50 € unter Ausblendung weiterer Erwägungen auf das Strafmaß hat.

Um allerdings auf diese Weise Strafgerechtigkeit herzustellen, kommt es auf *einen* Umstand in besonderer Weise an: Der Vergleichsfall darf kein willkürlich seitens eines einzelnen Strafrichters gewählter sein. Dies betrifft zum einen die vertikale Strafgerechtigkeit im Sinne der Gleichbeurteilung sachlich gleich gelagerter Fälle *durch diesen Richter*. Er darf daher nicht in jeder neuen Strafzumessungsentscheidung einen anderen Vergleichsfall wählen, den er zum Ausgangspunkt seiner Überlegungen nimmt.⁴³ Ebenso kann horizontale Strafgerechtigkeit allein unter der Voraussetzung hergestellt werden, dass eine gewisse

³³ Freund, GA 1999, 509 (536).

³⁴ Siehe schon Freund, GA 1999, 509 (536 m. Fn. 100) dazu, dass die beste Akzeptanz Erwartung allein die in der Mitte liegende Strafe, nicht etwa eine Strafe am unteren Rand der Schuldangemessenheit (so aber Grasnich, JA 1990, 81 [87]) aufweist. Dabei kennzeichnet die „Mitte“, dass es um die Streubreite faktisch anzutreffender Einzelfallbewertungen geht. Weil die „goldene Mitte“ am ehesten dem entspricht, was auch vom Verurteilten und der Gemeinschaft als „richtig“ akzeptiert wird, ist eine Orientierung an Extremen zu vermeiden.

³⁵ So auch Meier, Sanktionen (Fn. 8), S. 238, 257.

³⁶ *BGH*, StV 1986, 57 (57); ebenso BayObLG, JR 2002, 166 (167); vgl. *BGH*, StV 1995, 173 (173 f.) und Meier, Sanktionen (Fn. 8), S. 241. Ähnlich auch Streng, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 120, 201; Miebach/Maier, in: MüKo, 3. Aufl. (2016), § 46 Rn. 90. Ausführlich zur Kontrolle der Üblichkeit durch den *BGH* Maurer, Komparative Strafzumessung, 2005, S. 139 ff.

³⁷ *BGH*, NJW 1979, 1666 (1667); einschränkend *BGH*, NJW 2011, 2597 (2599).

³⁸ *BGH*, NStZ-RR 1997, 196 (197); NJW 1979, 1666 (1667); ebenso Meier, Sanktionen (Fn. 8), S. 238.

³⁹ *BGH*, NJW 2011, 2597 (2599); NStZ-RR 1997, 196 (197); Urt. v. 7.1.1992 – 5 StR 614/91; NJW 1979, 1666 (1667 f.).

⁴⁰ *BGH*, NJW 1973, 1805 (1806). Gegen eine Orientierung am praktisch häufigsten Fall auch *BGH*, NJW 1976, 2355 (2355). Auch Meier, Sanktionen (Fn. 8), S. 241 bejaht die fortbestehende Notwendigkeit der Berücksichtigung der individuellen Umstände des jeweiligen Falls.

⁴¹ Zum Maßstab siehe Miebach/Maier, in: MüKo, § 46 Rn. 84, 302, 314.

⁴² Allgemein zum Problem sprachlicher Unschärfe für Vergleiche Maurer, Strafzumessung (Fn. 36), S. 181.

⁴³ Siehe dazu, dass es sich hierbei um einen Verstoß gegen den Gleichheitssatz des Art. 3 GG handelte, Sarstedt, in: Verhandlungen des 41. Deutschen Juristentages, II, 1956, D 29, 32.

Übereinstimmung im Hinblick darauf besteht, was reichsspezifisch als „Leitfall“ eingestuft werden kann. Hierbei handelt es sich um besonders häufig auftretende Fälle, die in den strafzumessungsrelevanten Eigenschaften übereinstimmen. Damit sind zum Beispiel auch innerhalb der Diebstahlsfälle nach § 242 StGB bestimmte Erscheinungsformen als Untergruppen zu bilden. Nach normativer Absicherung der Differenzierung zwischen solchen „Leitfällen“ kann dann mit den Untergruppen weitergearbeitet werden. Vereinfacht gesprochen, müssen sich sämtliche Strafrichter *in Deutschland* in einem ersten Schritt an einem gemeinsamen Fall eines bestimmten Deliktstyps als Einstieg für ihre Strafzumessungserwägungen orientieren. An dieser Stelle zeigen sich freilich die eigentlichen Probleme der Umsetzung einer komparativen Strafzumessung und damit der Realisierung von mehr Strafgerechtigkeit.

1. Bestimmung des strafzumessungsrechtlichen Einstiegsfalls anhand rechtlich anerkannter gesellschaftlicher Wertvorstellungen

Die Gewichtung individuellen Verhaltensunrechts geht mit einer besonderen Herausforderung für den zur Rechtskonkretisierung Aufgerufenen einher. An diesem Punkt seiner rechtlichen Beurteilung des Falles kommt er nicht umhin, eine Komponente einzubeziehen, die die Bedeutung von Strafurteilen in ihrer Funktion der (Wieder-)Herstellung des Rechtsfriedens nach einer begangenen Straftat für den davon Betroffenen sowie die rechtlich verfasste Gemeinschaft in spezifischer Weise unterstreicht. Wieviel individuelles Unrecht wiegt, hängt auch von den rechtlich anerkannten Bewertungsmaßstäben derjenigen Gesellschaft ab, in der es sich ereignet hat.⁴⁴ Der Strafrichter muss daher innerhalb der Strafzumessungsentcheidung den Blick auf die allgemeinen Wertvorstellungen weiten, soweit diese rechtlich akzeptabel sind. Umso besser er als Seismograph im Hinblick auf eine solche Gewichtung von Verhaltensunrecht auftritt, desto mehr gelingt durch sein Urteil die (Wieder-)Herstellung des Rechtsfriedens durch Akzeptanz seiner Entscheidung.

Dabei sei zur Klarstellung betont, dass es an dieser Stelle nicht darum gehen darf, die Bestrafung des Täters in den Zweck zu stellen, Effekte bei Dritten im Sinne der Abschreckung oder Erziehung herbeizuführen.⁴⁵ Anderenfalls verkäme der Einzelne zum bloßen Mittel zum Zweck, was nicht in Einklang steht mit seiner verfassungsrechtlich garantierten Menschenwürde. Wenn hier also die Rede davon ist, dass der Richter die Bewertung des individuellen Fehlverhaltens von den rechtlich aner-

kennenswerten gesellschaftlichen Wertvorstellungen abhängig machen muss, ist damit nicht gemeint, dass auf einen bestimmten Zustand innerhalb der Rechtsgemeinschaft in spezifischer Weise – entgegenwirkend – reagiert werden soll. Mit der Gewichtung des individuellen Fehlverhaltens hat es nichts zu tun, wenn sich beispielsweise ähnliche Taten in der jüngeren Vergangenheit häufen, sodass eine besonders harte Bestrafung unter Umständen den Effekt zeitigen könnte, die übrigen Gesellschaftsmitglieder daran zu erinnern, wie wichtig die Normeinhaltung (auch) in diesem Bereich ist. (Richtige) Strafzumessung darf nicht mit der Gestaltung gesellschaftlicher Verhältnisse verwechselt werden.

Was allerdings (nicht zuletzt)⁴⁶ innerhalb der Strafzumessung durchaus von Bedeutung ist, sind Veränderungen im Hinblick auf allgemeine Wertvorstellungen: ob beispielsweise der Angriff auf die Ehre einer Person durch eine bestimmte Äußerung nach wie vor ebenso schwerwiegend eingestuft wird oder nicht. Diese Einschätzungen können sich mit der Zeit verändern, worauf auch der Richter zu reagieren und seine Bewertung entsprechend anzupassen hat. Komparative Strafzumessung bedeutet vor diesem Hintergrund keine Festschreibung eines bestimmten Zustandes.⁴⁷ Die Beurteilung, dass es sich bei einer spezifischen Tatbegehungsvariante um den strafzumessungsrechtlichen Einstiegsfall handelt, ist nicht in Stein gemeißelt. Sie ist ihrerseits der Anpassung an künftige Entwicklungen zugänglich, die durch allgemein geänderte, rechtlich anerkannte Bewertungsmaßstäbe ausgelöst werden kann. Vor diesem Hintergrund kommt auch eine blinde Übernahme fremder Erwägungen nicht in Betracht.⁴⁸ Der Vergleich mit anderen Strafmaßen kann immer nur eine (kritische) Orientierungshilfe für die eigene Entscheidung, niemals aber deren Determinierung sein.

2. Defizite empirischer Erkenntnisse zum strafzumessungsrechtlichen Einstiegsfall

In der Theorie lässt sich daher durchaus eine Methode benennen, die eine größere Strafgerechtigkeit in die deutsche Strafzumessungspraxis hineinbringen kann. Verkürzt lässt sich dies wie folgt darstellen: Der Strafrichter muss diejenigen Faktoren zunächst abstrakt benennen, die bei einer spezifischen Deliktsbegehung für das Tatgewicht eine Rolle spielen. Er hat dann zu analysieren, wie solche Fälle durch andere Gerichte bewertet werden und kann daraus ableiten, was verbreitet als angemessen empfunden wird. So kann er die Erkenntnis erlangen, welche Tatbegehungsvariante durchschnittlich als reichsspezifisch typischer Fall der Deliktsverwirklichung eingestuft und besonders häufig in einer bestimmten Weise geahndet

⁴⁴ Freund, GA 1999, 509 (536).

⁴⁵ Zur Dysfunktionalität entsprechender Strafzwecke innerhalb eines Tatstrafrechts siehe Rostalski, Tatbegriff (Fn. 14), S. 36 ff., 49 ff.

⁴⁶ Geänderte gesellschaftliche Wertvorstellungen können, soweit sie rechtliche Anerkennung verdienen, auch in anderen Bereichen des Strafrechts Bedeutung erlangen. Zu denken ist allein an die Ebene der Strafbegründung: Die Strafbarkeit des Ehebruchs (§ 172 StGB i.d.F. bis zum 1.9.1969) wird von den meisten derzeit als unangemessene Reaktion auf ein Verhalten angesehen, das zumindest moralisch von vielen nach wie vor verurteilt wird. Die Aufhebung der Strafbarkeit deckt sich mit dieser veränderten gesellschaftlichen Einschätzung.

⁴⁷ Siehe zu dieser Befürchtung aber Maurer, Strafzumessung (Fn. 36), S. 182.

⁴⁸ Maurer, Strafzumessung (Fn. 36), S. 182 spricht von der Notwendigkeit „revisionsfeste[r] Sanktionsalternativen“. Siehe ferner die Nachweise in Fn. 39.

wird. Für die Beurteilung seines eigenen Falles kann er sich nunmehr hieran orientieren. Dabei gilt es insbesondere, die Spezifika seines Entscheidungsgegenstandes herauszuarbeiten und zu bestimmen, ob es sich dabei um Gründe handelt, die ein Abweichen nach oben oder nach unten ausgehend von dem strafzumessungsrechtlichen Einstiegsfall rechtfertigen.

In der Praxis fehlt es den Strafrichtern allerdings in großem Umfang an der Möglichkeit, sich Zugang zu entsprechenden empirischen Erkenntnissen zu verschaffen.⁴⁹ Die gängige Praxis an deutschen Gerichten, zumindest gewisse „Haustarife“ festzulegen,⁵⁰ kann als Ausdruck des Fehlens empirischer Daten über den eigenen Einflussradius hinaus gewertet werden. Sie sollte insofern auch nicht vorschnell verurteilt werden. Der Sache nach handelt es sich dabei um den Versuch, jedenfalls in einer gewissen Reichweite, horizontale Strafgerechtigkeit herzustellen.⁵¹ Dass die regionalen „Haustarife“ unter Umständen nicht mit den Üblichkeiten in anderen Regionen übereinstimmen, ist vor diesem Hintergrund keine böse Absicht, sondern lediglich die Folge des Umstands, dass entsprechende Einsichten für den einzelnen Richter kaum zu erzielen sind.

Gründe dafür, weshalb es an entsprechenden Erkenntnismöglichkeiten fehlt, sind vielschichtig. In erster Linie hängt dies damit zusammen, dass eine Erfassung der Entscheidungen von Ausgangsgerichten in den gängigen juristischen Datenbanken in der Breite nicht erfolgt. Es ist daher für einen Hamburger Richter nicht ohne Weiteres durch einen Blick in juris oder beck-online möglich, zu prüfen, wie ein Münchener Gericht in einem Fall des einfachen Diebstahls einer Sache im Wert von 50 Euro entschieden hat. Wollte er sich hierüber Klarheit verschaffen, bedürfte es gesteigerter Mühen, die – die entsprechende Bereitschaft des Einzelnen unterstellt – nicht selten bereits an den föderalen Strukturen und der chronischen Überlastung deutscher Gerichte scheitern dürften.

Dabei steht eines fest: Wäre es möglich, eine große Zahl an Urteilen zu betrachten, so ließen sich regionale Unter-

schiede eliminieren. Aussagekräftig wird ein solcher Vergleich aber erst, wenn die jeweiligen Besonderheiten der betrachteten Fälle berücksichtigt werden.⁵² So ließe sich mittels statistischer Verfahren der Einfluss der jeweils enthaltenen Strafzumessungserwägungen auf das Strafmaß isolieren.⁵³ Ein Vergleich ist dann auch zwischen Fällen möglich, die nicht in allen, sondern nur in manchen Details übereinstimmen. Insofern ist *Meier* zu widersprechen, der annimmt, eine größere Vergleichsmasse erfordere Abstriche bei der Vergleichbarkeit.⁵⁴ Das Gegenteil ist der Fall.⁵⁵ Die eigentliche Schwierigkeit liegt vielmehr in der Sammlung und Klassifizierung einer hinreichenden Menge an Vergleichsmaterial.⁵⁶

V. Transparenz gegen die Willkür: Datenbanklösung

In der Gerichtspraxis finden sich (zumindest mancherorts) interne Strafmaßstabellen („Haustarife“), die für häufig vorkommende Delikte (wie Diebstahl, einfache Körperverletzung, Verkehrsdelikte) und bestimmte Konstellationen die am Gerichtsort üblichen Taxen oder konkretisierten Strafraumen enthalten. Teilweise sind diese sogar verschriftlicht.⁵⁷ Sie spiegeln die Erwartungen der Richterkollegen an eine „richtige“ Entscheidung wider und erzeugen einen gewissen sozialen Konformitätsdruck.⁵⁸ Im Ergebnis kann dies zur Vereinheitlichung der Strafzumessung in der jeweiligen Region führen.⁵⁹

Insofern enthält die Praxis schon jetzt komparative Elemente. Diese sind jedoch regional begrenzt.⁶⁰ Eine überregionale Berücksichtigung ist nicht möglich, da die „Tarife“ nicht öffentlich bekannt sind.⁶¹ Notwendig ist nach dem Gesagten eine umfassende und transparente empirische Grundlage. Hierzu ist mehrfach vorgeschlagen worden, eine bundesweite Datenbank zur Sammlung aller Strafzumessungsentscheidungen zu errichten.⁶² Darin sollen die Urteile entsprechend der enthaltenen Zumessungserwägungen klassifiziert werden, sodass anhand der Merkmale des zu entscheidenden Falls die statistische Verteilung des Strafmaßes aufgrund zahlreicher möglichst passender Vergleichsentscheidungen gefunden werden kann.⁶³ Hieraus soll sich nur eine Orientierungshilfe

⁴⁹ *Grasnack*, JA 1990, 81 (85).

⁵⁰ Dazu noch unten V.

⁵¹ *Grasnack*, JA 1990, 81 (81).

⁵² Zwar gelingt *Grundies*, in: Hermann/Pöge, S. 295 ff. die Berücksichtigung einer sehr großen Datenmenge von gut 1,5 Millionen Entscheidungen (S. 298). Die Auswertung beschränkt sich aber auf die Daten des BZR (S. 297). Ein solches Vorgehen hat aufgrund der damit einhergehenden Unmöglichkeit, die jeweiligen Einzelfallumstände hinreichend detailliert zu berücksichtigen (vgl. S. 299 f.), nicht die hier erforderliche Aussagekraft. Vgl. zur Bedeutung der Berücksichtigung normativer Unterschiede zwischen den untersuchten Fällen für die Beurteilung regionaler Strafmaßvarianz (insb. bei kleineren Vergleichsgruppen) *Albrecht*, Strafzumessung (Fn. 9), S. 348 ff.

⁵³ Als Beispiel mag hier die Untersuchung von *Albrecht*, Strafzumessung (Fn. 9), S. 612 ff. dienen.

⁵⁴ *Meier*, Sanktionen (Fn. 8), S. 238.

⁵⁵ Vgl. auch *Maurer*, Strafzumessung (Fn. 36), S. 182.

⁵⁶ Vgl. *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 120a; kritisch zur oft unzureichenden empirischen Vergleichsbasis auch *Verrel*, JR 2002, 168 (168). *BGH*, NJW 2011, 2597 (2599) spricht angesichts der praktischen Schwierigkeiten von einer Überforderung der Gerichte durch eine Pflicht zum Vergleich mit anderen Urteilen.

⁵⁷ *Kudlich/Koch*, NJW 2018, 2762 (2763); *Meier*, Sanktionen (Fn. 8), S. 240. Ein beispielhafter Abdruck aus dem Bereich der Verkehrsdelikte findet sich bei *Schäfer/Sander/van Gemmeren*, Praxis der Strafzumessung, 6. Aufl. (2017), Rn. 1720 ff.

⁵⁸ Vgl. *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 120. Diese Besonderheit interner Richtlinien übersieht *Maurer*, Strafzumessung (Fn. 36), S. 177.

⁵⁹ Zur Bedeutung informeller „Tarife“ für die Strafzumessungsentscheidung *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 119 m.w.N.

⁶⁰ *Meier*, Sanktionen (Fn. 8), S. 240; *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 119; vgl. bereits *Dreher*, MDR 1961, 343 (344). Zu den Variationen in der Strafzumessung siehe oben unter II.

⁶¹ *Meier*, Sanktionen (Fn. 8), S. 240; kritisch dazu (bezogen auf interne Richtlinien der StA) *Kinzig*, in: Schönke/Schröder, § 46 Rn. 72.

⁶² Entsprechende Forderungen finden sich bei *Kaspar*, Sentencing Guidelines (Fn. 6), C 115; *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 202.

⁶³ *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 202.

ergeben, nicht aber die Einzelfallentscheidung unter Berücksichtigung der Besonderheiten des jeweiligen Falls ersetzt werden.⁶⁴

Das Problem besteht auch hier in der Datengrundlage. Bisher ist eine solche nicht in hinreichendem Umfang vorhanden – sieht man einmal von den oben erwähnten internen Richtlinien ab.⁶⁵ Die Erstellung wäre wiederum extrem aufwendig, da sämtliche Urteile des für maßgeblich erachteten Zeitraums erfasst und händisch entsprechend den in den Texten enthaltenen Erwägungen klassifiziert werden müssten. Dies könnte für neue Urteile zwar unmittelbar durch die entscheidenden Gerichte geschehen, würde aber auch insofern eine Mehrbelastung bei ohnehin schon angespannter Personalsituation bedeuten.

VI. Legal-Tech-basierte (automatisiert erstellte) Datenbanklösung

Die Lösung für dieses Problem besteht in der automatisierten Erstellung der Datenbank mittels Legal Tech.⁶⁶ Dazu ist eine Software nötig, die nach Eingabe der Urteile die darin enthaltenen Strafzumessungserwägungen erkennt, mit den hierauf beruhenden Strafmaßen in Verbindung bringt und die Relation aus beidem in eine Datenbank schreibt. Diese Datenbank kann dann zur Abfrage detaillierter Statistiken aufgrund der jeweiligen Zumessungskriterien genutzt werden, wobei mittels statistischer Verfahren bei hinreichend großer Datenmenge auch Zusammenhänge zwischen Zumessungserwägungen gefunden werden können, die sich nicht unmittelbar aus dem Vergleich zweier Urteile ergeben.⁶⁷

Wesentliche technische Schwierigkeit ist dabei die Erkennung der Zumessungserwägungen durch das Programm. Denn selbst die fortschrittlichste „künstliche Intelligenz“ kann (zum jetzigen Zeitpunkt) nicht wirklich *verstehen*, worum es in einem Text geht.⁶⁸ Jedoch haben sich KI-Systeme als außerordentlich nützlich gezeigt, wenn es um das Auffinden von Korrelationen und Mustern in großen Datenmengen geht.⁶⁹ Hier hilft die relativ formale und gleichförmige Struktur von Strafurteilen weiter. Zwar macht § 267 StPO keine detaillierten Vorgaben für den Aufbau eines Urteils. In der Praxis ist es aber üblich, die Zumessungserwägungen gebündelt in einem eigenen Abschnitt zu benennen und näher darzulegen. Dabei werden häufig sehr ähnliche Formulierungen verwendet, wie etwa

„zu Gunsten des Angeklagten [wurde] berücksichtigt, dass“, „gegen den Angeklagten spricht“ usw.⁷⁰ Die Variationen in den Formulierungen lassen sich mittels Natural Language Processing (NLP) erfassen.⁷¹ Die Unterscheidung, welche Angaben im Urteil zur Strafzumessung gehören und welche nicht, erfolgt anhand des Kontextes solcher Formulierungen.

Bedeutsam für die Entnahme der Zumessungserwägungen aus dem Text der Urteilsbegründung ist naturgemäß, dass sie dort *vollständig* aufgezählt sind. § 267 Abs. 3 S. 1 Hs. 2 StPO enthält jedoch eine Beschränkung auf die „bestimmend[en]“ Erwägungen. Nicht alles, was in irgendeiner Form bei der Strafzumessung berücksichtigt worden ist, muss sich also zwingend auch im Urteil wiederfinden.⁷² Beim Vergleich mehrerer Urteile ist aber ohnehin eine gewisse Pauschalisierung und Fokussierung auf wichtige Faktoren notwendig. Die Berücksichtigung von Feinheiten kann nur anhand der Umstände des konkreten Falls erfolgen. Angesichts der überragenden empirischen Bedeutung einiger weniger Faktoren⁷³ erscheint dies auch hinnehmbar. Schwerer wiegt eine Abkürzung der Urteilsgründe gemäß § 267 Abs. 4 StPO – hier können die Strafzumessungserwägungen ganz oder teilweise entfallen. Derartige Urteile können für eine vergleichende Strafzumessung daher nicht oder allenfalls eingeschränkt berücksichtigt werden.⁷⁴

In der Praxis hat sich gezeigt, dass die Begründungen, die im Urteilstext angegeben werden, in Wahrheit nicht unbedingt auch (den erwarteten) Einfluss auf die Strafhöhe haben. So ist ein Teil der Erwägungen offenbar gänzlich wirkungslos, manche eigentlich strafmildernde Angaben korrelieren sogar mit erhöhten Strafen.⁷⁵ Auch derartige Befunde sprechen aber nicht gegen ein vergleichendes Vorgehen, solange sie im Rahmen dieser Methode aufgedeckt werden. Hier zeigt sich erneut, wie wichtig die Einbeziehung großer Datenmengen ist. Aussagekräftig bleiben die Ergebnisse trotzdem, da die im Urteil enthaltenen Erwägungen „im Kern“ tragend sind.⁷⁶

Neben den Zumessungserwägungen und dem Strafmaß sollte die Datenbank auch die zugehörigen Urteile selbst enthalten. So können die Anwender die gefundenen Ergebnisse überprüfen und insbesondere nach Besonderheiten gegenüber dem ihnen vorliegenden Fall suchen. In

⁶⁴ *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 202.

⁶⁵ Vgl. zum parallelen Problem bei der Schaffung einer Strafzumessungskommission *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 120a.

⁶⁶ „Legal Tech“ ist die Kurzform für „Legal Technology“, also die Unterstützung juristischer Tätigkeiten durch automatisierte Informationsverarbeitung, siehe dazu *Groh*, in: Weber, Creifelds kompakt, Rechtswörterbuch, 2019 (Stichwort „Legal Tech“); ähnlich *Wagner*, Legal Tech und Legal Robots, 2018, S. 1, dort auch zu verschiedenen Klassifikationen (S. 6 ff.).

⁶⁷ Vgl. dazu bereits oben unter IV.2.

⁶⁸ Vgl. zu „Information Retrieval“-Systemen *Wagner*, Legal Tech (Fn. 66), S. 24, aber auch S. 33 f. zur angeblich abweichenden Konzeption bei IBMs „Watson“. Insofern ähnlich *Frese*, NJW 2015, 2090 (2091). Jedoch stellt auch Watson letztlich nur Zusammenhänge her, ohne den tatsächlichen Sinn einer Aussage zu erfassen, vgl. *High*, The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works, S. 5.

⁶⁹ Siehe z.B. *Lu/Li/Chen/Kim/Serikawa*, Mobile Netw Appl 23 (2018), 368 (369); *Hopgood*, Computer 36 (2003), 24 (26).

⁷⁰ Die Zitate entstammen dem Urteil des *LG Verden* vom 18.3.2010 – 1 KLs 17/09.

⁷¹ Zum Einsatz von NLP bei juristischen Texten siehe etwa *Chalkidis/Kampas*, Artificial Intelligence and Law 27 (2019), 171.

⁷² *Miebach/Maier*, in: MüKo, § 46 Rn. 84, 302, 314.

⁷³ Laut *Meier*, Sanktionen (Fn. 8), S. 256 und *Kinzig*, in: Schönke/Schröder, § 46 Rn. 72 (jeweils m.w.N.) hängt die Strafhöhe statistisch insbesondere von der Tatschwere (etwa der Schadenshöhe beim Diebstahl) und den Vorstrafen des Täters ab.

⁷⁴ *BGH*, NJW 2011, 2597 (2598).

⁷⁵ *Albrecht*, ZStW 102 (1990), 596 (620 ff.). Angesichts der Divergenz von geschriebenen Erwägungen und tatsächlicher Strafhöhenermittlung spricht *Maurer*, Strafzumessung (Fn. 36), S. 179 von der Strafzumessungstheorie als „reine[r] Begründungswissenschaft“.

⁷⁶ *Albrecht*, ZStW 102 (1990), 596 (626).

Anlehnung an *Streng* folgt hieraus ein zweistufiges Vorgehen: Zunächst wird anhand besonders wichtiger und universalisierbarer Merkmale nach dem üblichen Strafmaß in insofern vergleichbaren Fällen gesucht, wobei das System im besten Fall die Möglichkeit bietet, den Einfluss der angegebenen Merkmale zu isolieren. In einem zweiten Schritt erfolgt dann eine Individualisierung der Strafe anhand der beim Vergleich nicht berücksichtigten Merkmale des konkreten Falls und der Unterschiede zu den Vergleichsfällen.⁷⁷

Hier besteht nun der Vorteil, dass die digitale Unterstützung die Auswertung größerer Datenmengen ermöglicht. Folglich ist keine zu starke Beschränkung auf wenige Vergleichsmerkmale nötig. Gleichwohl muss sich auch dieses System auf quantifizierbare Merkmale fokussieren.⁷⁸ Das können neben konkreten Zahlenwerten (etwa die Schadenshöhe beim Betrug) auch binäre Variablen wie die Existenz eines Geständnisses, einschlägiger Vorstrafen oder von Verfahrensverzögerungen sein, wobei jeweils auch feinere Abstufungen (Glaubhaftigkeit und Reue, Dauer und Ursache von Verzögerungen etc.) möglich sind. Nur schwer erfassbar werden aber auf absehbare Zeit komplexere Erwägungen etwa bezüglich einer besonderen kriminellen Energie, eines Mitwirkens des Opfers oder Ähnlichem sein. Insofern zeigt sich die besondere Bedeutung der Individualisierung der Strafe auf *Strengs* zweiter Stufe, da solche nicht quantifizierbaren Merkmale in bestimmten Situationen andere Erwägungen sogar deutlich überwiegen können. So wird auf diese Weise letztlich auch der Forderung des *BGH* Rechnung getragen, beim Vergleich mit anderen Urteilen tatsächlich sämtliche Umstände zu berücksichtigen.⁷⁹ Eine unkritische Übernahme fremder Überlegungen zu einer anderen Tat darf auch bei komparativem Vorgehen nicht erfolgen.⁸⁰

Soll die Datenbank Urteile im Volltext enthalten, müssen diese anonymisiert werden (vgl. Art. 10 S. 1 DSGVO⁸¹). Damit müssen zumindest der Name und sonstige Daten jedenfalls des Angeklagten entfernt werden, aufgrund derer dieser eindeutig identifiziert werden kann. Um die Identifikation auch unter Zuhilfenahme weiterer Informationen zu vermeiden, wird es außerdem notwendig sein, Orts- und Zeitangaben sowie Informationen zu Zeugen, Mittätern etc. zu schwärzen.⁸² Auch hier kann Legal Tech die Arbeit wesentlich erleichtern, denn auch Daten dieser

Art können aufgrund ihres Kontextes durch KI erkannt und automatisiert aus den Urteilen entfernt werden.

Die Vorteile einer in dieser Form erstellten Datenbank für Gerichte und sonstige Rechtsanwender dürften auf der Hand liegen. Vergleiche mit anderen Urteilen sind ein ohnehin genutztes Mittel, es fehlt jedoch eine zuverlässige Datengrundlage.⁸³ Gleiches gilt aber auch für die Kontrolle der Strafzumessung durch die Revisionsgerichte,⁸⁴ die bei der Ermittlung des Üblichen bisher überwiegend nur auf ihr subjektives Empfinden und ihre eigenen Erfahrungen rekurrieren (können).⁸⁵ Hinzu kommt das Interesse von Staatsanwaltschaften und Strafverteidigern an Transparenz hinsichtlich der zu erwartenden Strafen.⁸⁶ Insbesondere für die Verteidigung dürfte vielfach die Prozessstrategie davon abhängen, wie sich welche Erwägungen durchschnittlich auswirken – gerade auch, wenn diese Auswirkungen regional variieren. Schließlich ist auch das wissenschaftliche Interesse an einer flächendeckenden Erhebung über die Unterschiede und Gemeinsamkeiten bei der Strafzumessung nicht zu unterschätzen.

Problematisch ist derzeit noch die Verfügbarkeit von Rohdaten. Insbesondere Algorithmen, die auf Verfahren des maschinellen Lernens basieren, benötigen einen großen Pool an Trainingsdaten. Frei verfügbare Urteile gibt es zwar in großer Zahl, jedoch stammen diese meist nur von Obergerichten und enthalten damit keine oder nicht die vollständigen Strafzumessungserwägungen. Spätestens für eine flächendeckende und repräsentative Erhebung sind ohnehin Urteile aller deutschen Gerichte notwendig, ohne dass diese in irgendeiner Form gefiltert werden dürfen.

VII. Schluss

Es gibt in Deutschland erhebliche Abweichungen der Strafzumessungsentscheidungen, die sich nicht durchweg durch sachliche Unterschiede in den zugrundeliegenden Fällen rechtfertigen lassen. Vielmehr spielen sowohl die individuelle Richterpersönlichkeit als auch der Entscheidungsort eine nicht unerhebliche Rolle für das konkrete Strafmaß. Dabei gilt es als offenes Geheimnis, dass innerhalb mancher Richterkollegien interne Orientierungshilfen verwendet werden, die zu einer einheitlicheren Straf-

⁷⁷ Vgl. *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 120; zustimmend *Meier*, Sanktionen (Fn. 8), S. 238 f.

⁷⁸ Zu einer solchen Beschränkung i.R.e. Datenbanklösung auch *Kaspar*, Sentencing Guidelines (Fn. 6), C 115.

⁷⁹ Vgl. *BGH*, NJW 2011, 2597 (2598).

⁸⁰ Vgl. die Nachweise in Fn. 39.

⁸¹ Zum hieraus grundsätzlich folgenden Behördenvorbehalt für nicht anonymisierte Daten über strafrechtliche Verurteilungen siehe *Bäcker*, in: BeckOK, 28. Ed. (2019), Art. 10 DSGVO Rn. 7 m.w.N. auch zur Gegenauffassung.

⁸² Angesichts der hohen Individualität von Strafurteilen besteht aber u.U. weiter die Möglichkeit, die geschwärzten Urteile den Originalen zuzuordnen (unterstellt, die Originale sind verfügbar). Ob daher tatsächlich von einer Anonymisierung oder lediglich von einer Pseudonymisierung gesprochen werden kann und ob eine solche den Anforderungen der DSGVO genügt, kann hier nicht entschieden werden. Zu den hohen Anforderungen an die Anonymisierung personenbezogener Daten siehe *Karg*, in: Simitis/Hornung/Spiecker genannt Döhmman, 2019, Art. 4 Nr. 1 DSGVO Rn. 57 ff.

⁸³ Siehe oben unter V.

⁸⁴ *Streng*, in: Kindhäuser/Neumann/Paeffgen, § 46 Rn. 202; *Kaspar*, Sentencing Guidelines (Fn. 6), C 115.

⁸⁵ Vgl. etwa *BGH*, StV 1995, 173 (173 f.); StV 1986, 57 (57); als Gegenbeispiel darf *BayObLG*, JR 2002, 166 (167) gelten. Zum Ganzen *Verrel*, JR 2002, 168 (168).

⁸⁶ Auch *Kinzig*, in: Schönke/Schröder, § 46 Rn. 72 m.w.N. fordert die Bekanntgabe der internen Zumessungsmaßstäbe gegenüber den übrigen Verfahrensbeteiligten, bezieht sich dabei allerdings auf die StA.

zumessung führen sollen. Der Vergleich mit anderen Tatbegehungsvarianten ausgehend von einem bereichsspezifisch typischen Einstiegsfall ist als Methode zur Herstellung von Strafgerechtigkeit grundsätzlich vorzugswürdig. Er setzt aber voraus, dass es einen *gemeinsamen* Fixpunkt gibt, der sämtliche Strafzumessungsentscheidungen bestimmt – und nicht bloß diejenigen in einzelnen Regionen des Landes. Gegenwärtig lässt sich dies allerdings aufgrund der fehlenden Transparenz der regionalen „Haustarife“ sowie des eingeschränkten Zugangs zu den Entscheidungen der Ausgangsinstanzen kaum umsetzen. Hier kann eine umfassende Datenbank Abhilfe schaffen. Diese händisch zu erstellen, erscheint angesichts der knappen Ressourcen der Justiz nahezu ausgeschlossen. Dank der zunehmenden Möglichkeiten der Informationstechnolo-

gie könnte dieser Schritt in Zukunft aber entfallen. Automatisierte Analysemethoden sind heute in der Lage, auch komplexe Zusammenhänge aus Texten in natürlicher Sprache zu extrahieren. Mittels auf diese Weise gewonnener Daten ließe sich die Intersubjektivität der Strafzumessungserwägungen ebenso verbessern wie ihre Transparenz. *Dreher* kann daher im Hinblick auf seine Kritik am geheimen „Metermaß“ des Richters darin zugestimmt werden, dass Geheimnistuerei in diesem Bereich besonders fehl am Platze ist. Einen „Fixpunkt“ muss es aber geben – und zwar einen gemeinsamen. Um diesen zu finden, bedarf es entsprechender Forschung, die nicht zuletzt der hohen Bedeutung der Strafgerechtigkeit für die Akzeptanz von Strafurteilen Rechnung trägt.⁸⁷

⁸⁷ Siehe zu einem solchen Projekt <http://legaltechcologne.de/strafzumessung-mit-legaltech-transparent-machen-smart-sentencing-data-base>; https://www.rostalski.jura.uni-koeln.de/sites/strafrechtprof2/Dokumente/Smart_Sentencing_Pressemitteilung_des_Lgal_Tech_Lab_Cologne.pdf (zuletzt abgerufen am 3.9.2019).