Criminal Law Forum (2017) 28:437–475
DOI 10.1007/s10609-017-9327-6

WOLFGANG FRISCH*

FROM DISPARITY IN SENTENCING TOWARDS SENTENCING EQUALITY: THE GERMAN EXPERIENCE**

ABSTRACT. This article addresses movements in Germany towards greater uniformity in sentencing since the entry into force of the first federal criminal code in 1871, with a particular focus on developments since the second half of the twentieth century. The author reviews the empirical evidence for sentence disparity; addresses the limitations of the constitutional principle of equality as a means for overcoming sentencing disparity; identifies the causes of sentence disparity and tracks practical efforts made to overcome that disparity, specifically through the introduction of prosecutorial guidelines and judicial manuals for certain frequent types of offending. He also explains and critiques different theoretical models for increasing sentencing equality, and reviews the advantages and potential uses of sentencing information systems such as a judgment database created in Japan. The author then turns to the contribution made by the appellate courts to greater uniformity in sentencing and expresses the hope that recent trends towards intensified guidance for trial courts in appeals concerning tax offences will be extended to other types of offences as well.

I SENTENCING DISPARITY AS EMPIRICAL FACT AND PROBLEM

1.1 *Critiques of Sentencing Disparity in German Sentencing Practice*

Concerns about sentencing disparity arose soon after the first federal criminal code, the *Reichsstrafgesetzbuch (RStGB)* (which had replaced the diverse criminal codes of the German federal states, and had been expected to bring about legal unity in the area of criminal law) had entered into force in 1871. As early as 1874, R. Medem claimed that when it came to sentencing, "uncertainty and disparity"[1]

* Professor of Criminal Law, Criminal Procedure and Legal Theory Emeritus at the University of Freiburg, Freiburg, Germany. E-mail: wolfgang.frisch@jura.uni-freiburg.de

** Translated by Antje du Bois-Pedain.

[1] Medem, "Über Strafzumessung und Strafmaß" (1874) 26 *Der Gerichtssaal* 590, 591.

were to be found instead of uniform application of the criminal code. As is shown by the frequency with which such complaints were repeated in the decades which followed, disparity in the apportionment of sentences was not a transitional phenomenon on the way to a soon-to-be-achieved uniformity but rather a permanent state of affairs in criminal sentencing practice. In 1906, Wilhelm Kahl observed "appalling disparities" indicative of a failure of the "sentencing apparatus"; in 1921, Hermann Mannheim criticised "insufferable inequality of punishments".[2] Others put the point in more extreme language, chiding the "anarchic state" of criminal sentencing.[3] In the 1950s, Hellmuth von Weber, a criminal law professor at the University of Bonn, spoke of the "chaos" of sentencing practice,[4] and the founder of modern sentencing theory in Germany, Hans-Jürgen Bruns, complained in 1974 – along with a number of other criminal law scholars – of the existence of a "state of extreme divergence in sentencing practice" which was "in the longer term intolerable".[5] Such comments were not only made by criminal law theorists, who may be thought to have a natural tendency to find fault with practice. The chorus of critical voices was joined by practitioners, too. Werner Sarstedt, a well-known judge at the *Bundesgerichtshof (BGH)* (the Federal Court of Justice, which is highest court of appeal in the German court system in civil as well as certain criminal matters), in 1955 spoke of up to four-fold differences in sentence length for comparable crimes and was concerned about the compatibility of this state of affairs with the constitutional guarantee of equality.[6] Other judges likewise referred to "frightful discrepancies" and complained about "sentencing confusion", "the miserable state of sentencing jurisprudence" and even an "emergency in sentencing".[7]

[2] Kahl, "Reform der Strafzumessung" (1906) 11 *Deutsche Juristen-Zeitung* 895; Mannheim, "Über Gleichmäßigkeit und Systematik in der richterlichen Strafzumessung" (1921) 54 *Zeitschrift für die gesamte Strafrechtswissenschaft* 40, 41.

[3] Drost, *Das Ermessen des Strafrichters* (Cologne: Heymanns, 1930) p. 117.

[4] Von Weber, *Die richterliche Strafzumessung* (Karlsruhe: Müller, 1956) p. 19.

[5] Bruns, *Strafzumessungsrecht: Gesamtdarstellung* (Cologne: Heymanns, 1974) p. 507.

[6] Sarstedt, "In welcher Weise empfiehlt es sich, die Grenzen des strafrichterlichen Ermessens… zu regeln…?", in *Verhandlungen des 41. Deutschen Juristentags*, vol. 2 (1955), D 30, 39.

[7] Middendorf, "Das Maß des Richters. Eine vorläufige Bilanz der Diskussion um die Strafzumessung" (1970) 7 *Blutalkohol* 257, 286; Seib, "Gleichmäßigkeit des Strafens, ein Prüfstein der Gerechtigkeit" (1971) 8 *Blutalkohol* 18; Tröndle, "Das Problem der Strafzumessungsempfehlungen" (1971) 8 *Blutalkohol* 73, 74.

Admittedly, these drastic descriptions of sentencing practice were formulated some time ago. But it stands to reason that, if it was indeed correct to diagnose intolerable sentencing discrepancy over the course of more than 100 years of sentencing under a uniform federal criminal code, it would be unlikely for the problem to have disappeared completely by now. Some improvements may, of course, have come to pass – how and why, is an issue to which I will return later in this article. For now, the question remains whether the quoted observations were accurate descriptions of sentencing practice and the sentence patters resulting from it. Could these criticisms not, perhaps, stem from random, unrepresentative or deceptive impressions which make sentencing practice appear far worse than it really was (and is)? What are the empirical bases for these assessments?

1.2 *The Empirical Foundations of the Claimed Disparity in Sentencing*

The assessments sketched above, and similar views expressed today, are supported by a range of rather different observations, including some systematic research. Given this broad and varied basis for these complaints, it is plausible to conclude that they were at their core well founded (and may well still be legitimate to some extent today).

Judges and other practitioners quoted above who have weighed in on this debate have usually relied on their personal experience and observation rather than on any systematic study.[8] A good example is Sarstedt's remark that, in his capacity as judge at the *BGH* (which hears criminal appeals against judgments in the more serious criminal cases), "he sat on appeals against the judgments of two neighbouring trial court panels, one of which in general used to give the accused about four times as heavy a punishment as he could have expected to receive if his case had been tried before the panel next door".[9] It is likely that similar observations form the basis of critical comments by other practitioners about the state of sentencing practice. It is of course possible to object to the evidential value of such observations and statements on the basis that one has no way of ascertaining whether the cases that were given such different sentences were really sufficiently similar. But the impressions and observations of experienced practitioners, who over many years on the bench have developed a good eye for what matters when it comes to passing sentence,

[8] Streng, *Strafzumessung und relative Gerechtigkeit* (Heidelberg: v.Decker, 1984) p. 5 f.

[9] Sarstedt, fn. 6 above, D 39.

are certainly not completely without value. This is reinforced by the fact that these observations by practitioners are supported by a number of scientific studies and investigations.

Of particular significance in this regard are studies of sentencing outcomes in identical cases that were decided by different judicial bodies, or even by the same judicial body again. The latter could (until 1965) occur when the trial court's judgment was set aside on appeal and the case was sent back to the same judge, who then had to rehear the case and resentence the accused.[10] It can also happen when the same case is accidentally decided twice by the same judge. Regarding the first type of case, Wilhelm Haddenhorst published a follow-up study of a cohort of 76 successful appeals on a point of procedure decided in 1964, where the appellant had subsequently been retried and resentenced by the same judge or panel. His data showed that, even when the trial court again arrived at the same findings as to facts and guilt at the conclusion of the retrial, the punishments imposed on the convicted offender the second time round were invariably distinctly less severe.[11] Even more striking are the discoveries made by Karl Peters in the context of his research into sources of error in the criminal trial. He came across 40 cases where the same person had accidentally been convicted and sentenced twice for the same crime – albeit usually by different judges or panels. The punishments imposed on the two occasions were rarely the same; in most cases there were differences. In the most extreme cases, the sentence on one occasion was six times as severe as the sentence imposed on the other sentencing occasion.[12]

It stands to reason that such discrepancies not only arise on the rare occasions where the same case is tried and sentenced on two different occasions, but are in fact much more pervasive. This is demonstrated by studies in which the complete identity of cases, which is so rare in real life, is artificially created by providing sen-

[10] Before the reform of criminal procedure by the *Strafprozessänderungsgesetz* (StPÄG) of 19 December 1964 took effect, an appellate court could, upon allowing the appeal, direct that the case be sent back to the original trial judge to be retried and resentenced before this judge. This is no longer permitted. Now, unless the appellate court is able to render a final judgment, the case must be sent back to a different judge or panel of the same court. See section 354 (2) of the Code of Criminal Procedure (*Strafprozessordnung, StPO*).

[11] Haddenhorst, *Die Einwirkung der Verfahrensrüge auf die tatrichterlichen Feststellungen im Strafverfahren* (Karlsruhe: Müller, 1971) pp. 82, 85.

[12] Peters, "Praxis der Strafzumessung und Sanktionen" in Göppinger and Hartmann (eds), *Kriminologische Gegenwartsfragen*, Vol. 10 (Stuttgart: Enke, 1972) pp. 51, 60; on Peters' studies, see also Streng, fn. 8 above, p. 10 f.

tencers with fictitious cases for sentencing. In one such study by Franz Streng, about 500 judges, prosecutors and legal trainees in Lower Saxony were asked to propose sentences in six different hypothetical cases.[13] The result, given the discussion above, should not be considered surprising: there was very significant discrepancy in the sanctions suggested by the study participants which in the less serious cases ranged from monetary fines to imprisonment of up to two years (not in all cases suspended); and when the case vignettes contained serious offences such as rape or manslaughter, the suggested duration of imprisonment diverged by multiples of years. Comparable studies have found similar results.[14] If one bears in mind that the suggested sentences reflected the punishments thought appropriate by criminal justice practitioners and legally trained persons soon about to join the profession, one can expect such discrepancies also to manifest themselves in actual cases.

A weighty indication of significant disparities in sentencing is, finally, provided by statistical analyses of the sentencing practice of German courts. Some studies of this kind were conducted already in the first half on the twentieth century by Otto Woerner and by Franz Exner[15]; others date from the second half of the twentieth century.[16] These studies analyse and compare the sentencing patterns for selected crimes in certain court districts. Technically speaking, these studies do not concern identical, but different cases. But if, based on a uniform practice of case allocation, the courts decide a sufficiently large number of cases within a certain category, the differences between the individual cases balance each other out, such that significant discrepancies in the sanctions imposed are important indicators

[13] Streng, fn. 8 above, pp. 75 ff., 78 ff., 95 ff., 102 ff.

[14] See e.g. Opp and Peukert, *Ideologie und Fakten in der Rechtsprechung* (München: Goldmann, 1971); Hood, *Sentencing the Motoring Offender* (London: Heinemann, 1972) p. 143; D. Peters, *Richter im Dienst der Macht* (Stuttgart: Enke, 1973); Streng, fn. 8 above, p. 59 ff. (with further references).

[15] Compare *Exner*, *Studien über die Strafzumessungspraxis der deutschen Gerichte* (Leipzig: Wiegand, 1931) p. 46 ff. (on regional differences); Woerner, *Die Frage der Gleichmäßigkeit der Strafzumessung im deutschen Reich* (Munich: Beck, 1907) pp. 23, 53. Exner's study was the first study based on a systematic evaluation of national statistical data on judgments rendered in different regional courts.

[16] See e.g. Schiel, *Unterschiede in der deutschen Strafrechtsprechung* (Hamburg: Kriminalistik Verlag, 1963) pp. 23 ff., 55 f., 81 ff.; Maurer, *Komparative Strafzumessung* (Berlin: Duncker & Humblot, 2005) p. 25 ff.

of actual disparity.[17] Statistical analyses conducted by Exner in the 1930s found sentence discrepancies of up to a multiple of four; Hans-Jörg Albrecht in 1980 discovered for the region Baden-Württemberg that courts in the southern districts of the region had handed down small fines (below 100 German marks) three times more frequently than courts in the northern districts; the inverse was true for large fines (above 2000 German marks) which courts in the north had handed down five times as frequently as courts in the south.[18] The income patterns of persons living in the north and in the south could not explain this north–south differential. Moreover, the discrepancy persisted even after the day fines system was introduced (which responds to differences in income when setting the sum per day, whereas the severity of the sentence is expressed in the number of day fines imposed).

Significant differences were found not only in respect of the quantity of punishment. They also emerged in respect of the question whether sentences of imprisonment were suspended or not. A study commissioned by the Ministry of Justice of the regional state Baden-Württemberg found, in respect of convictions for drink-driving, that in some local courts only 1.4 per cent of the sentences imposed for this offence were suspended, whereas in other local courts more than 80 per cent were suspended.[19] A comparison of the rates at which sentences of imprisonment generally were suspended between 1964 and 1966 by trial courts located in the neighbouring appeal court districts Stuttgart and Karlsruhe found that this took place at twice the rate in the latter district compared to the former. Other studies, such as one based on a review of court files by Heinz Schöch,[20] suggested that a similar situation obtained in the rest of the country as well.

1.3 *Qualifications and Open Questions*

The studies cited date predominantly from the 1960s and the 1970s; for this period of time they support the drastic observations quoted in the introduction. It is difficult to know to what extent they also reflect

[17] See Streng, fn. 8 above, p. 5; for a critical assessment Hörnle, *Tatproportionale Strafzumessung* (Berlin: Duncker & Humblot, 1999) p. 66.

[18] H.-J. Albrecht, *Strafzumessung und Vollstreckung bei Geldstrafen* (Berlin: Duncker & Humblot, 1980) pp. 88, 206 f.

[19] See Streng, fn. 8 above, p. 7 f. (with further references).

[20] See Schöch, "Möglichkeiten und Grenzen einer Typisierung der Strafzumessung bei Verkehrsdelikten mit Hilfe empirischer Methoden" in Göppinger and Hartmann (eds), fn. 12 above, pp. 128, 133.

the present state of sentencing practice. More recent studies (done around the year 2000) have failed to find similarly large discrepancies in sentencing.[21] This development can plausibly be attributed to successful efforts made by the prosecutorial agencies and the regional ministries of justice to bring about greater uniformity, particularly with regard to the disposition of crimes committed very frequently and with little variation in the mode of commission, and in the approach taken to suspension of sentences of imprisonment. But one should not conclude from this too easily that sentencing equality has been achieved across the board: we still come across difficult-to-explain individual sentencing decisions (particularly in some cases that lead to successful appeals), astonishing divergences between the sentencing outcomes for different participants in the same crime, and surprising differences in sentencing outcomes when cases are resentenced after a retrial.[22] The persistence of sentencing discrepancy is made all the more plausible by the fact that the causes of discrepancy – which I will address in greater detail below – have not all been effectively removed. The extent of the remaining uncertainties and resulting discrepancies becomes apparent when we move beyond routine, frequent cases to newly created offences, to offences which span a wide variety of conduct and contexts, and to newly recognised mitigating or aggravating factors such as giving credit to the defendant for his willingness to made a full or partial confession, thus allowing the case to be concluded more easily. This last point provides a particularly powerful example. Notwithstanding the absence of a system of "guilty pleas" in the formal sense, German trial courts have traditionally rewarded a defendant's admission of the offence charged with a more lenient sentence. Over time, an informal practice of negotiating sentencing outcomes between defence lawyer, prosecutor and judge (referred to as "deals") developed on the back of this, and was eventually put on a statutory footing (section 257c *StPO*, inserted into the Code of Criminal Procedure in 2009). When the Federal Constitutional Court had to rule on the constitutionality of this statutory provision, the court commissioned an empirical study

[21] See e.g. H.J. Albrecht, *Strafzumessung bei schwerer Kriminalität* (Berlin: Duncker & Humblot, 1994) p. 492 (concluding that sentencing practice for serious crimes is largely consistent); similarly Hoppenworth, *Strafzumessung beim Raub* (Munich: VVF Verlag, 1991) pp. 23 ff., 54 ff., 258 ff., 266. For further empirical data see Maurer, fn. 16 above, pp. 31 ff., 37 ff.

[22] On this point, compare also Maurer, fn. 16 above, p. 43 ff. and Hörnle, fn. 17 above, p. 67 f.

into trial court practice of settling criminal cases through a "deal". This study, conducted by Karsten Altenhain,[23] found such dramatic differences in how different courts acknowledged cooperation by the defendant that the presentation of the study's findings caused great consternation among those present in the oral hearing before the Constitutional Court.

This means that despite the improvements brought about since the 1960s and 1970s there remains a significant need to further reduce sentencing disparity. The reason why one cannot just sit back and accept these disparities is easily stated. Such disparities are incompatible with the constitutional rights to equal treatment and to justice, and moreover undermine the legitimacy of the criminal justice system in the public eye.[24] They are serious impediments to positive general prevention by reinforcing respect for the law, because they send inconsistent and irrational signals. They can also make the reintegration of offenders more difficult.

So, the question is: what can be done to make sentencing more uniform?

II THE PRINCIPLE OF EQUALITY – AN INSTRUMENT FOR OVERCOMING SENTENCING DISPARITY?

Given that the principle of equality (enshrined in Article 3 of the German Constitution) obliges the state to treat like cases alike, one might be tempted to turn to this principle – which after all appears to be the obvious principle that is violated by disparity in sentencing – to make sentencing more uniform and just.

This was certainly the view taken by some convicted offenders who felt that they had been treated unequally and unjustly, in comparison to other offenders convicted of comparable crimes, by the sentences that were imposed upon them. They claimed that the much heavier punishments that had been imposed upon them, as compared to the punishments given by other courts in cases of this sort, violated their constitutional right to equality. Such arguments have been advanced before appellate courts and also on occasion by way of a constitu-

[23] Altenhain, Dietmeyer and May, *Die Praxis der Absprache im Strafprozess* (Baden-Baden: Nomos, 2013) pp. 116 f., 182, 183; see further Frisch, "Schuldprinzip und Absprachen" (forthcoming in *Festschrift für Franz Streng*). The Constitutional Court issued its judgment on 19 March 2013 (BVerfGE 133, 168).

[24] Streng, fn. 8 above, p. 13 ff.; see further Frisch, fn. 23 above and BGHSt 28, 318 ff., 324 f.

tional complaint. They are made with particular frequency in cases where different parties to the same crime are given distinctly different punishments for what the offenders consider their broadly equal participation in the underlying offence – whether the sentences are passed in the same trial by the same court, or by different courts on different sentencing occasions.[25]

So far, neither the appellate courts nor the Constitutional Court have shown themselves impressed by this line of argument. The Constitutional Court informed the complainant that the only obligation imposed by the principle of equality on an individual trial court was "to pass judgment without making distinctions based on any prejudice against the accused, and not to make any arbitrary distinctions in determining the appropriate punishment".[26] No violation of the principle of equality would arise simply because the trial court – when determining the punishment in accordance with this standard – had arrived at a punishment that differed from what another court had done in a comparable case. While the result – the rejection of the complaint – seems ultimately correct, the reasoning leaves something to be desired since one can hardly deny the intuitive appeal of the thought that divergent sentences in comparable cases are incompatible with the principle of equality.[27]

The true reason why one cannot show a violation of the principle of equality in an individual case is a different one. The individual has a right to have his case decided according to law and, in this sense, correctly.[28] He has no right, as such, to have his case decided in the same way as another case – since that other case may well have been incorrectly decided or may be problematic in some other respect. Provided that his case has been decided in accordance with the law, the convicted offender cannot show that he is in any way unfairly burdened by how his case has been decided. Consequently his complaint must fail. The fact that another individual may have received

[25] See e.g. BGHSt 28, 318 ff., 324; BGH StV 1981, 122 f.; BGH StV 1987, 435 f.; BGH StV 1995, 557; BGH StV 2009, 244 f.; BGH NJW 2011, 2597 ff. = JR 2012, 249 ff. noted by Streng ibid.; further examples in Streng, *Strafrechtliche Sanktionen*, 3rd edn (Stuttgart: Kohlhammer, 2012) marginal note 664; Maurer, fn. 16 above, p. 152 ff.

[26] See e.g. BVerfGE 1, 332 ff., 345.

[27] Similarly Streng, fn. 8 above, p. 15; see further Warda, *Dogmatische Grundlagen des richterlichen Ermessens im Strafrecht* (Cologne: Heymanns, 1962) p. 157 ff.

[28] This is pointed out inter alia by Bruns, fn. 5 above, p. 507; Warda, fn. 27 above, p. 157 f.; and BGHSt 28, 318, 324.

more favourable treatment, possibly in violation of the law, imposes no burden on the complainant. He has no right to benefit equally from unlawfulness.[29] Moreover, since judges are obliged to pass judgment in accordance with their understanding of the law, they cannot be expected to change their decision based on the opinions of other courts which they do not share or consider mistaken.[30] The situation is, however, different when the same judge has without satisfactory explanation imposed vastly different sentences on co-perpetrators whose respective contributions to the crime were essentially the same. In such a case, an appeal should and likely will succeed on the grounds that the reasons given fail to provide a satisfactory explanation for the differential treatment.[31]

Thus, the principle of equality is in practice not much help against disparity in sentencing. Individual sentencing decisions will, for the reasons given, usually not be challengeable as violations of the principle of equality. As for the requirement to ensure sentencing equality and to overcome unwarranted disparity at the systemic level, the diagnosis of disparity unfortunately does not come with a recipe for its cure. To find a cure, we must look at the causes of sentencing disparity. Insight into the causes of disparity may provide us with a strategy for bringing about greater sentencing equality by addressing these causes.

III THE CAUSES OF SENTENCE DISPARITY

The causes of sentence disparity have been investigated by a number of recent criminological and sociological studies. The main drivers of disparity were, however, already identified very soon after the first federal criminal code came into force in Germany in 1871. Broadly, they can be found in the combined influence of two phenomena addressed in the next sub-section.

[29] From the Constitutional Court's jurisprudence, see e.g. BVerfGE 50, 142, 166; as the Federal Court of Appeal in Administrative Matters (*Bundesverwaltungsgericht*) pointed out, any argument based on parallel cases that were not decided in accordance with the law is therefore irrelevant (BVerwGE 34, 269, 283). For further references and discussion, see e.g. Jarass and Pieroth, *Grundgesetz. Kommentar*, 14th edition (Munich: Beck, 2016), commentary to Art. 3, marginal note 36.

[30] BGHSt 28, 318, 324 f.; BGH NJW 2011, 2597 ff. and the literature cited in fn. 25 above.

[31] See again the cases and literature cited in fn. 25 above.

3.1 *The Absence of Objective Directives Regarding the Appropriate Measure of Punishment is Compensated for by Personal Views*

The first of these phenomena was already described by Medem, one of the first critics of disparity in sentencing after the entry into force of the German criminal code. Penal theory has – he said in 1874 – "hardly any idea by what measure the punishment should be apportioned to the crime".[32] This was taken up by Franz von Liszt when he said that, in the absence of the law providing a measuring rod, apportioning the punishment is "like groping about in the dark".[33] Von Liszt and Adolf Wach also remarked that what tends to become the determining factors in groping about for the right sentence are, then, "the coincidences that have resulted in a particular composition of the panel of the court, the personal opinions and inspirations of the judges, their temperament and their digestion."[34]

Modern criminological studies avoid such crude and unscientific generalisations in pinpointing the sources of sentencing discrepancy. But the gist of their findings is remarkably similar. One can see this clearly in the results of the already mentioned large-scale study conducted by Streng, in which more than 500 judges, prosecutors and legal trainees in Lower Saxony were given six fictitious cases to sentence. Streng's analysis of the discrepancies in the suggested punishments indicates that important determining factors were: the different views held by the study participants regarding the causes of crime, in particular whether they were committed to classic freedom-based conceptions of criminal guilt or to sociological explanations for crime; which penal aims they found most important; what their role was in the criminal process (judge, prosecutor, legal trainee); and which court they were based at.[35] Other studies have attempted to demonstrate correlations between certain socio-demographic factors of the decision-maker (such as age, gender, religious belief and social origins) and the severity of the supported punishments.[36]

---

[32] Medem, fn. 1 above, pp. 591, 603.

[33] Von Liszt, "Kriminalpolitische Aufgaben (1889–1892)" in von Liszt, *Strafrechtliche Aufsätze und Vorträge*, Vol. 1 (Berlin: Guttentag, 1905) pp. 290, 393.

[34] Wach, *Die Reform der Freiheitsstrafe* (Leipzig: Duncker and Humblot, 1890) p. 41; von Liszt, "Die Reform der Freiheitsstrafe (1890)" in von Liszt, fn. 33 above, pp. 511, 533.

[35] Streng, fn. 8 above, pp. 75 ff., 80 ff., 94 ff., 107 ff., 132 ff., 138 ff., 156 ff., 160 ff., 186 ff., 189 ff., 207 ff., 209 ff.; see also Streng, fn. 25 above, marginal note 488 ff.

[36] See e.g. Opp and Peukert, fn. 14 above; Haisch, "Zur Anwendung der Attributionstheorie auf die Strafzumessung in simulierten Verfahren" (1980) 24 *Zeitschrift für die Anwendungsgebiete der Psychologie* 13; see further Streng, fn. 25 above, marginal note 488 ff.

One potential source of sentencing disparity – that individual decision-makers have different personal views of what the "right" punishment is – is therefore still in place. Of course, this cause can only have an effect if there is room for it to influence sentencing outcomes. If the law gave the decision-maker strict and binding instructions on how to decide particular cases, her personal attitudes and preferences could have no, or only a very small, effect on sentencing outcomes. One may also question whether the second source of sentencing disparity identified by Medem, von Liszt und Wach – the absence of guidance by penal theory regarding the measure of punishment – still presents the same problem as in earlier times. In contrast to previous times, we now have some explicit legislative guidance regarding the criteria to be applied in sentencing decisions, including what may or may not be taken into account in such decisions. Legislation also now addresses with much greater precision than in earlier times questions such as when a sentence ought to be suspended.[37] There is also a significant scholarly literature concerned with questions of sentencing.[38] And the jurisprudence of the appellate courts has played an important part in ensuring greater uniformity – as is evidenced by the fact that more than a third of successful appeals, are appeals against sentence.[39]

While all of this is true, it remains the case that the situation has in one important point changed little with respect to the measure by which punishment should be apportioned to achieve sentencing equality. To see this, we need to take a closer look at the content of the legislative provisions, the theoretical models on sentencing developed in the scholarly literature and the decisions of the appellate courts.

---

[37] Suspension of sentence was first regulated in sentencing law in 1953, when the predecessor provisions to what are now sections 56 ff. *StGB* were inserted into the German Criminal Code. Until then suspension of sentence was conceptualised as an act of grace. On the present regime of suspension of sentence, see further S Harrendorf, "Sentencing Thresholds in German Law and Practice: Legal and Empirical Aspects" (in this issue), Sect. 2.3.

[38] See e.g. Bruns, fn. 5 above; Schäfer, Sander and van Gemmeren, *Praxis der Strafzumessung*, 5th edn. (Munich: Beck, 2012); Streng, fn. 25 above, marginal note 479 ff.

[39] Statistics are given by Nack, "Aufhebungspraxis der Strafsenate des BGH" (1997) 17 *Neue Zeitschrift für Strafrecht* 153, 156 f.

3.2 *Legislative Reform has Failed to Address the Real Deficits*

The central provision concerning the type and measure of punishment (section 46 sub.-s. (1) of the German Criminal Code) merely states that the offender's personal guilt is the basis for sentencing, adding that the effects which the sentence can be expected to have on the offender's future life in society shall be taken into account. While this is more than nothing, it does not tell us all that much. The extent to which such effects should be taken into account remains an open question, as does the question whether and to what extent aspects of (positive or negative) general prevention may influence the sentence imposed. It is thus unsurprising that much is disputed even at the level of principle,[40] which opens the door for personal preferences for particular sentencing aims to be given effect.

Other central provisions, such as section 47 of the German Criminal Code which governs the selection of the type of sanction, and section 56 of the German Criminal Code which concerns the suspension of sentences of imprisonment, are somewhat more precise but still rely at crucial points on vague normative concepts wide open to divergent interpretations and concretisations, for example "whether the imposition of a prison sentence is necessary to defend the legal order" (section 47 sub.-s. (1)), or whether "special circumstances can be found" which permit the court "after a comprehensive evaluation of the offence and character of the offender" to suspend a sentence of between one and two years of imprisonment (section 56 sub.-s. (2)).[41]

Greater direction is provided by section 46 sub.-s. (2) and (3) of the German Criminal Code, which set out respectively circumstances which should, and circumstances which may not, be taken into account in apportioning the punishment (although in respect of the former, it is not stated whether the circumstances should operate as aggravating or as mitigating factors). Of course, many further questions arise after the relevant circumstances in the individual case

[40] On the debate concerning the significance of the principle stated in section 46 sub.-s. (1) StGB, see Stree and Kinzig, "§ 46 StGB" in Schönke and Schröder (eds), *Strafgesetzbuch. Kommentar*, 29th edn (Munich: Beck, 2014) marginal note 3 ff.; Streng, fn. 25 above, marginal note 521 ff.; Frisch, "Strafkonzept, Strafzumessungstatsachen und Maßstäbe der Strafzumessung in der Rechtsprechung des Bundesgerichtshofs" in Roxin and Widmaier (eds.), *50 Jahre Bundesgerichtshof*, vol. 4 (Munich: Beck, 2000) pp. 269, 270 ff.; Hörnle, fn. 17 above, p. 326 ff. (all with further references).

[41] See further, Schäfer, Sander and van Gemmeren, fn. 38 above, marginal note 155 ff.; Streng, fn. 25 above, marginal note 183 ff.

have been identified, and all the law has to say in this regard is that the sentencer ought to "balance these factors against each other" (section 46 sub.-s. (2), 1st sentence). How exactly this ought to be done and what the aim of the exercise should be, the law does not specify any further.

Particularly, the sentencer is given no further guidance regarding how he should move from having identified and balanced the relevant circumstances to a specific punishment. All that the law offers is a range within which the sentence must fall – say, from the lightest possible punishment[42] up to five years' imprisonment, or from three months to ten years of imprisonment. Within these very wide ranges the judge is instructed to identify the sentence which responds appropriately to the offender's guilt and takes the effects of the sentence on the future life of the offender adequately into account. Ultimately, this amounts to not much more legislative guidance today than when the Criminal Code of 1871 was passed.

3.3 *Questions Regarding the Determination of the Measure of Punishment are Insufficiently Resolved in the Scholarly Literature and the Appellate Jurisprudence*

Of course, today we have a developed academic literature on sentencing that did not exist when the Criminal Code was first enacted. But anyone who thinks that this literature provides clear standards that can guide the selection of sentence, will be disappointed.

The existing literature addresses – typically in conjunction with an analysis of relevant appeal court practice – a variety of practically important questions, such as: the purposes of punishment; the interpretation of the fundamental principle that the sentence should be based on the degree of the offender's personal guilt; which factors are relevant to sentence and when these factors count as aggravating or as mitigating; and the need to balance these factors. But very little is said about how to perform this balancing operation, its envisaged result, and by what means this result can be achieved, and nowhere is explained what balancing outcomes or what combination of factors should result in which concrete punishments. All one finds are references to the range of available punishments for the offence and the expectation that the concretely chosen sanction should reflect the gravity of the case under sentence.

[42] The shortest possible sentence of imprisonment is one month (section 38 sub-s. (2) StGB); the lowest possible fine is five day fines (section 40 sub.-s. (1), second sentence).

What the academic literature has to offer beyond the guidance provided by the statutory provisions themselves, is a particular conceptualisation of the range of available sanctions. This offers sentencers some signposts for the selection of an appropriate sentence. In essence, the literature says that the appropriate punishment cannot be located just anywhere on the permissible sentencing range, with the sentencer placing the given case somewhere on this range based on his personal opinion regarding the appropriate measure of punishment. This was how many had understood the task of the sentencing judge in the early years after the coming into force of the Criminal Code, and based on this notion it was then also maintained that a sentence that was within the permissible range of sentences could not be challenged on appeal on substantive grounds.[43] Today, however, the sentencing range is conceived of as a scale which ranges from the least serious conceivable case to the most serious conceivable case of the commission of this particular offence (the scale of relative seriousness of cases), which is matched by a parallel scale that starts with the minimum available punishment for the offence and ends with the most serious punishment that may be imposed for the offence.[44] In practice, this means that for cases that are classified as ''less serious'', ''medium-serious'', etc., the sentencer never has the full range of penalties permitted by the offence definition available to him. Rather, only a narrower band-width, essentially a ''range within the range'' suitable for that category of cases, remains open to him. For instance, the sentence for less serious cases must remain near the bottom end of the range; cases of medium seriousness may attract

[43] See e.g., RGSt 8, 76, 77 (no appeal lies against sentence if the sentence remains within the minimum and maximum terms authorised by law and the trial court's deliberations have not proceeded from any other error or law). For details, see Frisch, ''Die erweiterte Revision'', in Arnold et al (eds.), *Menschengerechtes Strafrecht. Festschrift für Albin Eser zum 70. Geburtstag* (Munich: Beck, 2005) pp. 257, 259 f.

[44] On this conception of the sentencing range, see e.g. BGHSt 27, 2 ff.; BGH StV 1983, 102 and 117; BGH StV 1984, 114; Bruns, fn. 5 above, p. 81 ff.; Bruns, fn. 31 above, p. 60 ff.; Dreher, *Über die gerechte Strafe* (Heidelberg: Schneider, 1947) p. 61 ff.; Dreher, ''Über Strafrahmen'' in Frisch and Schmid (eds.), *Festschrift für Hans-Jürgen Bruns zum 70. Geburtstag* (Cologne: Heymanns, 1978) pp. 141, 149 ff.; Frisch, *Revisionsrechtliche Probleme der Strafzumessung* (Cologne: Heymanns, 1971) p. 161 ff.; Meier, *Strafrechtliche Sanktionen*, 3rd edn. (Heidelberg: Springer, 2009) p. 207; Schäfer, Sander and van Gemmeren, fn. 38 above, marginal note 622 ff. For criticism of this conception, see Hettinger, *Das Doppelverwertungsverbot bei strafrahmenbildenden Umständen* (Berlin: Duncker & Humblot, 1982) pp. 129 ff., 149; Streng, fn. 8 above, p. 42 ff.

punishments that are significantly more serious than the lightest permissible sanction but still remain clearly within the lower half of the full range of punishments, etc. In consequence of this approach, the appellate courts have allowed appeals against sentence when a frequent, run-of-the-mill, not particularly serious instance of a particular offence was given a punishment located near the mid-point on the way to the maximum punishment.[45] The same would hold if the sanction in a very serious case remained at or near the minimum punishment for the crime.

It is undeniable that this conception of the available sentence range – which I believe is the correct approach[46] – can provide sentencers with very important initial guidance in individual cases. But it certainly does not provide a measure capable of guiding decision makers towards a concrete quantum of punishment.[47] Take the punishment for a "low-to-medium-serious case" for an offence whose legislative sentence range is imprisonment for no less than three months and for no more than ten years. Whether any particular "low-to-medium-serious case" deserves one, two, or three years, or any of the many possible punishments in between, cannot be answered based on such rough notions. What is more, one would also need implementable objective guidance regarding when a case is less serious, medium serious, etc. Another reason why the principle of scaling violations of a particular offence definition according to notions of relative seriousness does not generate very differentiated guidance is that the qualifiers according to which cases are categorised are necessarily quite vague (less serious, medium serious, very serious). Another problem with treating the legislative range of permitted punishments as a gravity scale is that many of these ranges are by now outdated – and particularly, that many of them set unrealistically high maxima.[48] Here the first duty of those charged with applying the law is to determine what constitutes an appropriate

[45] BGHSt 27, 2, 4 f.

[46] Compare Frisch, fn. 44 above, p. 161 ff.; Frisch, "Maßstäbe der Tatproportionalität und Veränderungen des Sanktionenniveaus" in Frisch, von Hirsch and Albrecht (eds.), *Tatproportionalität* (Heidelberg: Müller, 2003) pp. 155, 159 ff.

[47] See Schäfer, Sander and van Gemmeren, fn. 38above, marginal note 623; Frisch, fn. 46 above, p. 160 f.

[48] Dreher, "Über Strafrahmen", fn. 44 above, pp. 141, 150 ff.; Streng, fn. 8 above, p. 44 f.; see further Frisch, "Gegenwärtiger Stand und Zukunftsperspektiven der Strafzumessungsdogmatik, Teil II" (1987) 99 *Zeitschrift für die gesamte Strafrechtswissenschaft* 751, 789 ff.

upper limit. Moreover, one criminal case may allow for the application of multiple offence penalty scales. This gives rise to a host of further questions and requires the development of additional criteria not provided by the theory of the gravity scale as such.

The remaining serious deficits of the legislative and jurisprudentially developed sentencing guidance remain uncompensated in the practice of the courts, including at appellate level. In the early decades after the coming into force of the 1871 Criminal Code the appellate courts were careful not to give any indication as to what they considered the appropriate sentence to be in any individual case. Once legal reform had extended the grounds of appeal to the Federal Court of Justice (*BGH*) to permit appeals against sentence, judicial comment on the appropriate punishment became more frequent. But even now these pronouncements are limited to criticisms of the sentence that was in fact imposed. So far, appellate decisions have not really set out in general terms which sanctions are appropriate for which kinds of cases – although more recent judgments concerning the sanctioning regime for tax fraud and tax evasion appear to be an exception.[49]

3.4 *Summary: Why Inconsistency Persists*

As the discussion above has shown, we still lack legislative provisions, scholarly interpretations and case law that could provide sentencers with a clear identification of the right quantity of punishment. The exact placement of a case on the range of possible sanctions remains, of necessity, a choice influenced by the sentencer's personal views and the existence of other determinants. The decision is influenced, among other factors, by the sentencer's opinions on matters of criminal policy, his preference for particular sentencing aims, his basic inclination towards milder or sharper sanctions, and whether he happens to be aware of how other judges and courts have sentenced in similar cases. The absence of consistency in sentencing is, against this backdrop, only to be expected. If we want to bring about greater consistency – and the need for such greater consistency can, on the grounds shown above, not be doubted – this is where we must start. We must, first, provide sentencers with objective criteria which ensure that questions regarding the right quantity of punishment can be answered consistently. Secondly, we must ensure that these criteria are not just adequate to the task but also followed – and, if necessary, enforced.

[49] See section V.5.2. below.

Let us turn to the first of these questions: How can sentencers be provided with criteria for sentencing that ensure regularity and consistency in punishment? What kinds of criteria are suitable for this task?

IV POSSIBLE STRATEGIES FOR ENSURING GREATER CONSISTENCY IN SENTENCING

There has been no shortage of suggestions on how the quantification of punishment could be brought closer to the ideal of greater uniformity in sentencing. Some of these suggestions have been partially implemented and have indeed had some equalising effects. Other suggestions do not offer acceptable solutions and should not be pursued any further.

4.1 *Predictable and Uniform Quantification of Punishments Through Mathematical Models?*

The most ambitious approach to achieving consistency in sentencing involves the mathematisation of the sentencing exercise. Various scholars have attempted to develop such models for Germany.[50] In essence – and also in their simplest form[51] – these models determine the quantum of punishment by assigning defined numerical values to certain (aggravating or mitigating) factors, for use in conjunction with tables containing numerical scores that translate into specific punishments or punishment ranges. The numerical score of an individual case calculated under such a system then reflects the ''relative gravity'' of the case, and is matched to a concrete punishment. The method is similar to attempts that have been made to rationalise the evaluation of evidence by attaching numerical values to certain kinds

[50] Cf. *Bruckmann*, ''Vorschlag zur Reform des Strafzumessungsrechts'' (1973) 6 *Zeitschrift für Rechtspolitik* 30; Haag, *Rationale Strafzumessung* (Cologne: Heymanns, 1970); von Linstow, *Berechenbares Strafmaß* (Berlin: Schweitzer, 1974); Kohlschütter, *Die mathematische Modellierung der Strafzumessung* (Marburg: Tectum, 1998).

[51] The discussion in the text is oriented towards von Linstow's model which links the numerical values associated with different criteria through multiplication or addition adapted for certain offences, thus calculating a ''rough penal score'' (von Linstow, fn. 50 above, p. 10 f.). The application of further decision rules not related to the evaluation of the severity of the offending behaviour then results in a definite sanction (ibid, p. 27 ff.). On other models, see Frisch, fn. 46 above, p. 171 ff. and Maurer, fn. 16 above, p. 80 ff.

of evidence which are then used to calculate the likelihood of the truth of a proposition.[52] There are also parallels to statistical tables used for prognostic assessments.

Admittedly, such a method of sentence calculation could make sentencing outcomes more predictable and also more uniform, at least provided that it is differentiated enough and incorporates all relevant sentencing factors. But how operable is such a method in practice? And would its implementation without more generate acceptable sentencing outcomes?

The possibility of implementing this kind of sentencing model depends on the existence of generally acceptable criteria, according to which particular numerical values can be attached to certain sentencing factors, and on the basis of which particular aggregate numerical scores can then be associated with certain penalties. Unfortunately we lack such criteria, at least if we require the criteria to be generally accepted or at least generally acceptable.[53] If against this backdrop the numerical values assigned to different factors are simply set at a certain level, one risks generating sanctions that will not find widespread acceptance.[54] In order to avoid this and to generate acceptable punishments, one has to ensure that the aggregate numerical scores calculated under the model translate into a punishment that appears independently appropriate in the individual case (say, because it can be seen to be adequate in light of the seriousness of the offence and the culpability of the offender). If it is to be acceptable, the model cannot replace the standard that a punishment must be an appropriate response to the gravity of the offending behaviour but rather requires that the punishments it generates are demonstrably appropriate. So to lead to acceptable punishments this method requires extensive further work to ensure that the numerical score has been converted into an appropriate punishment. This is not a simplification of the sentencing process, but a complication – with only an apparent increase in rationality – that generates no advantages in the cases for which numerical tables have been developed.[55]

---

[52] Compare e.g. Ekelöf, "Beweiswürdigung, Beweislast und Anscheinsbeweis" (1962) 75 *Zeitschrift für Zivilprozess* 289, 292; Bruns, "Beweiswert" (1978) 91 *Zeitschrift für Zivilprozess* 647 ff.

[53] Streng, fn. 8 above, p. 314.

[54] Maurer, fn. 16 above, p. 87: "somewhat randomly assigned numerical values".

[55] See Frisch, fn. 46 above, pp. 155, 172; Streng, fn. 8 above, p. 314 f.; Maurer, fn. 16 above, pp. 80 ff., 88 f.

Now, it may be thought that this method would have advantages if the numerical scores are applied over time to many other cases – that it would ensure both consistency and acceptability of punishment on these subsequent occasions. But that would be an illusion. We have assumed, above, that the application of the table in a concrete case has been tested against the independently valid requirements for appropriate punishment. Other cases are different; we have no reason to believe that because the table generated an independently correct outcome in one particular set of factual circumstances it would also do so in another. We would have to engage in the same additional effort again. Moreover, even if the same score-generating sentencing factors reoccur in another case, we cannot know whether it is appropriate to score these factors the same, now that they appear in a new combination of circumstances. So to ensure the acceptability of the sentence one must always check the calculated result for its compliance with the standards which truly govern the appropriateness of a particular sentence.

In short: a system that replaces the traditional evaluative process sentencers engage in with mathematical calculations based on numerical scores, defined values, or assigned weights, is not defensible.[56] At best, such methods can aid a traditional sentencing exercise: if the "calculated" punishment were to be very different in severity from the sentence which the sentencer, having taken into account all the circumstances of the case, has in mind to impose, this can inspire a sentencer to reconsider the matter.[57] But, of course, it is not a foregone conclusion that it is then the calculated punishment which will prevail – it is rather more likely that the sentencer will proceed to impose the sentence that he thought appropriate in all the circumstances. Even so, it remains an advantage of the mathematical method that, on some occasions when it prompts a sentencer to rethink the punishment he was otherwise minded to impose, further reflection may well lead the sentencer to realise that his initial judgement was mistaken. Whether this rather small advantage justifies the enormous effort necessary to produce the system for calcu-

[56] Apart from the authors quoted in fn. 55, the following also agree with this assessment: Hassemer, "Die Formalisierung der Strafzumessungsentscheidung" (1978) 90 *Zeitschrift für die gesamte Strafrechtswissenschaft* 64, 76 f.; Köberer, *Iudex non calculat* (Frankfurt/Main: Lang, 1996) pp. 60 ff., 103 ff., 167 ff.; see further Frisch, fn. 46 above, p. 171 ff.

[57] For such a supporting role: Dubs, "Grundprobleme des Strafzumessungsrechts in der Schweiz" (1982) 94 *Zeitschrift für die gesamte Strafrechtswissenschaft* 161, 171 f.

lating sentences is more than doubtful – even more so since mistakes in relation to the overall evaluation of the appropriate punishment can be corrected in other ways. This also explains why a comprehensive system of sentence calculation in every individual case has so far not been developed.

4.2 *Ensuring Uniformity in Sentencing Through Relying on Notions of "Ordinary" or "Median" Cases?*

The efforts described above to create a mathematical method for sentencing are particular manifestations of attempts to achieve greater rationality and uniformity in sentencing through pre-evaluating sentencing factors and establishing rules for the correlation of these factors to sentencing outcomes. These attempts need to be contrasted with approaches that strive to increase rationality and uniformity in sentencing by orienting the sentencing decision around certain types of cases which function as "anchoring points" for the decision. Two types of mental constructs are in use: the "ordinary" case, and the "median" case.[58] The latter notion describes a hypothetical case which is imagined to fall half-way between the least and the most serious imaginable case that could arise under the provision, and the punishment associated with this case lies at the mid-point between the most and the least severe penalty authorised by the provision. By contrast, the "ordinary" case is the kind of case which statistically occurs most frequently. The experience-based ordinary case is, thus, not to be equated with the theoretical construct of the median case. With most criminal offence definitions, practical experience shows that the majority of cases falling under the provision are of relatively low gravity.[59] Hence the appropriate sanction for such ordinary cases is nowhere near the mid-point of authorised punishments, but must remain clearly beneath it – according to some au-

[58] For these definitions, see Bruns, fn. 5 above, p. 85 ff.; Bruns, *Neues Strafzumessungsrecht?* (Cologne: Heymanns, 1988) p. 63 ff.; Burns, "Die Bedeutung des Durchschnitts-, des Regel-, und des Normalfalles im Strafrecht" (1988) 43 *Juristenzeitung* 1053 ff.; Götting, *Gesetzliche Strafrahmen und Strafzumessungspraxis* (Frankfurt/Main: Lang, 1997) p. 60 ff., 213 ff.; Neumann, "Zur Bedeutung von Modellen in der Dogmatik des Strafzumessungsrechts" in Seebode (ed), Festschrift für Günter Spendel zum 70. Geburtstag (Berlin: De Gruyter, 1992) 435, 444 ff.; Ostermeyer, "Die Regelstrafe" (1966) 19 *Neue Juristische Wochenschrift* 2301; Schoene, "Regelstrafe" (1967) 20 *Neue Juristische Wochenschrift* 1118.

[59] BGHSt 27, 2, 4 f.; Bruns, fn. 5 above, p. 85; Ostermeyer, fn. 58 above, 2302.

thors, in the lower third of the authorised penalty range[60]; according to others, close to the minimum punishment.[61]

Reflection on what constitutes an ''ordinary'' and a ''median'' case for a particular kind of offence, and the penalties thought to be appropriate for either, is expected to assist sentencers not only in disposing of cases that are considered to reflect the features of an ordinary or of a median case, in that these notions enable uniform disposition of such cases. It is also expected to be useful for the disposition of unusual or outlier cases, in that it may help to clarify the degree to which the case at hand differs from the ordinary or median cases respectively, and thus to guide reflection on the degree of modification of sentence necessary to reflect the particular case's factual divergence from the ordinary, or the median, case.[62]

These attempts to focus on what punishments are appropriate in ordinary cases, and on clarifying – through reflection on the hypothetical median case – the hallmarks of the kind of case for which a punishment near the mid-point of authorised sanctions may be warranted, are certainly more realistic strategies for increasing the rationality and uniformity of sentencing practice than are efforts to attach defined and precise weights to certain sentencing factors. But even here, the gains are somewhat limited. A high degree of uniformity can only be expected to result in the ''ordinary'' and the ''median'' cases themselves. The more the case at hand diverges from these guiding concepts, the less guidance they are able to provide and the less one can expect to see comparable outcomes. And even the expectation of a uniform treatment of ordinary and median cases is at present not met. This is because we still lack generally accepted or even generally acceptable standards for what constitutes an ''ordinary'' and a ''median'' case with respect to the many different offences defined in the Criminal Code itself or in other legislation. These are addressed neither in academic commentaries nor in court judgments. To implement this approach, one would (at least for frequently applied offences) have to engage intensively in the identification of the ordinary and the median case, so as to develop a generally accepted system which provides the necessary orientation.

---

[60] See e.g. Meier, fn. 44 above, p. 209 f.; Ostermeyer, fn. 58 above, 2302. For empirical analysis, see Götting, fn. 58 above, p. 221 ff.

[61] See Bruns, fn. 5 above, p. 85.

[62] So Bruns, fn. 5 above, p. 85 f.; see already Sturm, ''Zur Lehre vom Strafmaß'' (1922) 34 *Zeitschrift für die gesamte Strafrechtswissenschaft* 64, 69.

How useful and profitable such an intense engagement with the identification of ordinary and median cases would be, is difficult to say. Realistically one could determine the features of the ordinary and the median case only on the basis of relatively comprehensive documentation and systematic analysis of the different contexts and ways in which certain crimes are committed.[63] At present we still lack such a comprehensive documentation. Available statistics merely provide information on punishments "by offence" but provide no detail as to the mode of commission of these offences and the punishments associated with different modes. But even if we had more detailed comprehensive documentation of the decided cases, an approach that concentrates entirely on the identification of ordinary and median cases and their respective sanctions in order to increase sentencing consistency, cannot be recommended. Such a method threatens to produce new, unproductive points of contention and is really a detour on the way to greater rationality and uniformity of sentencing decisions.

Assuming comprehensive detailed case documentation, the first new area of dispute is likely to be which of the documented cases represent ordinary or median cases respectively, and why. If the first point any commentator needs to settle is what constitutes an ordinary and a median case for a particular offence, this would tie down a commentator's limited resources of attention and argumentative capacity in a most unhelpful way. This is so because in order to achieve greater uniformity in sentencing, it is hardly necessary that one knows for each offence what an ordinary and a median case of its commission are. It is entirely sufficient to know which concrete sanctions have been, or would be, considered appropriate in a number of particular cases which fall under the offence definition – whether on the basis of decided and documented cases, or (as in some criminological studies[64]) based on hypothetical case scenarios. This information already supplies useful markers of orientation for a more uniform approach to sentencing these crimes. Since one knows which punishments have been, or would have been, considered appropriate in the previously evaluated cases, one can determine the punishment for the new case at hand by considering in what respects the case at

[63] To take each individual judge's impression of what, based on her previous experience, appears to be an ordinary case and sentence, as a guide for how to sentence future cases, is merely bound to cement existing divergences between different judges and courts.

[64] See Streng, fn. 8 above, pp. 64 ff., 75 ff.

hand differs from the previously evaluated cases, and what modification of the punishment appropriately reflects these differences. The wider the net has been cast to capture previously decided cases, and the greater the homogeneity of the previous case evaluations captured in that net, the higher the likelihood that new decisions guided by these prior evaluations will result in comparable outcomes. Hence, one does not really need to know or agree on which of the cases constitutes an ordinary case, and what are the features of a median case, in order to generate greater uniformity in sentencing. This also appears to be the unspoken basis for a number of other practical initiatives to increase uniformity in sentencing.

4.3 *Proposed Punishments for Certain Typical Cases: Guidelines, Etc.*

Disparity in sentencing has long been a matter of concern not just for academic commentators but also for practitioners, particularly when it comes to very frequently committed crimes whose mode of commission is highly similar (so-called mass crimes), and when the same questions are answered in contrary ways. Marked differences in the sentencing of mass crimes, and divergent answers given to obviously identical questions, do not only appear to indicate unequal treatment and a violation of distributive/comparative justice. They also undermine the reputation and public acceptance of the criminal justice system. Moreover they have damaging effects for positive general prevention through the reinforcement of the public's sense of right and wrong, since the messages they send out irritate rather than reaffirm respect for the law. Hence it is unsurprising that it has long been a concern of criminal justice practitioners to bring about greater uniformity in sentencing.

4.3.1 *Judicial Manuals for Road Traffic Offences; Administrative Guidelines for Prosecutors and Fiscal Authorities*

Sentencing for road traffic offences, particularly drink driving, leaving the scene of an accident without good cause, and offences of causing injury committed in the context of road traffic violations, has played a pioneering role in this regard. The high degree of similarity between the cases led to great irritation with divergent punishments and divergent answers to the question when sentences of imprisonment could be suspended or not; problems much discussed at practitioner congresses in the 1950s and 1960s. Judges present at these meetings developed manuals that recommended particular sanctions

(or narrow sanction ranges) for certain typical manifestations of such cases defined by a few important criteria (such as the degree of intoxication, the length of distance driven, etc.). These manuals also contained extensive additional discussion and guidance, beyond what was contained in the legislation itself, on the question when a sentence of imprisonment ought to be suspended, and particularly when such suspension of sentence was no longer appropriate.[65] Of course these manuals were not binding on a trial judge; they contained mere suggestions and advice which the trial judge was free to follow or not.

Somewhat more binding – albeit only indirectly – is the effect of guidelines issued by the regional states' general-directorates of prosecutors with a view to increasing uniformity in sentencing in certain areas.[66] These guidelines, which are based on resolutions of the assembly of regional justice ministers, are similar in structure and content to the sentencing manuals developed by judges for road traffic cases and also cover predominantly road traffic offences, but also some other crimes such as narcotics offences. Of course, these guidelines cannot bind the trial judge, given the constitutional guarantee of judicial independence; nor are they addressed to the trial judge. They are addressed to the prosecutors who deal with these cases, who are to base their suggested sentences in a written "penal order" procedure, and their submissions as to sentence in the oral hearing before the trial court, on these guidelines. As criminological research has shown, this gives the guidelines an indirect effect on judicial sentencing decisions. Research findings suggest that in areas without detailed statutory guidance, certain views expressed prior to the judicial decision tend to have significant effects on the decision. They function as anchoring points towards which the decision maker

[65] For detailed discussion, see Bruns, fn. 5 above, p. 187 ff.; Maurer, fn. 16 above, p. 175 ff.; Rastätter, *Die Strafzumessung bei der Steuerhinterziehung*, PhD thesis, University of Freiburg (2016), p. 132 ff.; Streng, fn. 8 above, p. 50 f.

[66] In Germany's federal structure, the administration of justice falls within the competence of the regional states. Hence, each regional state has its own justice minister and general-directorate of prosecutors. Nation-wide policies are agreed by the justice ministers in regular meetings and formalised in resolutions at these meetings. On prosecutorial guidelines see Bruns, fn. 5 above, p. 187; Rastätter, fn. 65 above, p. 132 ff. For examples of administrative guidance for prosecutors in respect of traffic offences, see further Schäfer, Sander and van Gemmeren, fn. 38 above, marginal note 941 ff.

orients his decision even if he does not completely follow the suggestion made.[67] The submission by the prosecutor as to sentence performs such an anchoring function, partly because the prosecutor is perceived as professionally competent and partly because the judge is aware that, if her decision deviates markedly from the sentence which the prosecution has suggested based on the guidelines, the prosecution might well appeal against sentence.

Of comparable significance to the prosecutorial guidelines which exist for some types of general criminality are fiscal administrative directives for fiscal crimes.[68] These directives likewise aim to ensure uniform treatment of tax investigations. Their structure and content is similar to the prosecution guidelines. They acquire significance for criminal trials because the tax investigators perform the role of the prosecuting authority in criminal tax prosecutions. It therefore falls to them to select the sanction when requesting a criminal court to issue a penal order, and to propose an appropriate punishment at the end of a criminal trial. Criminal prosecutors also defer to the tax authorities' directives in certain other cases where the tax authorities are interested parties. Moreover, recently the appeals courts have weighed in heavily on the disposal of criminal tax investigations and formulated certain adequacy requirements – a development to which I will return later.[69]

4.3.2 *Evaluating the Effects of the Guidelines and Bench Manuals*

Views differ on the efficacy of the efforts just described to increase consistency in sentencing. In this regard, the prosecution guidelines and judicial manuals share the fate of Anglo-American sentencing guidelines,[70] to which the former can be considered functionally equivalent in some areas of sentencing.

---

[67] See e.g. Englich and Mussweiler, "Sentencing Under Uncertainty. Anchoring Effects in the Courtroom" (2001) 31 *Journal of Applied Social Psychology* 1535; Streng, fn. 25 above, marginal note 498.

[68] See Meine, *Die Strafzumessung bei der Steuerhinterziehung* (Heidelberg: Müller, 1990) marginal note 120 ff.; further Rastätter, fn. 65 above, p. 272 ff.

[69] See section V. 5.2. below.

[70] In the German literature, these are discussed by Reichert, *Intersubjektivität durch Strafzumessungsrichtlinien* (Berlin: Duncker & Humblot, 1999) p. 166 ff.; Fischer, *Die Normierung der Strafzumessung nach Vorbild der U.S. Sentencing guidelines* (Baden-Baden: Nomos, 1999) pp. 119 ff., 128 ff.; see also Frisch, fn. 46 above, p. 167 f.; Maurer, fn. 16 above, p. 71 ff.; Streng, fn. 25 above, marginal note 764 ff.

Even those voices – both among academics and in the judiciary – who are sceptical of guidelines approaches[71] cannot deny that the prosecution guidelines have led to greater uniformity in sentencing in particular areas of criminality. The extreme discrepancies which could be observed in the 1950s and early 1960s in the sentencing of road traffic offences, and also more generally in respect of the decision to suspend or not to suspend sentences of imprisonment, have all but disappeared.[72] Given the anchoring function performed by the prosecution's submissions on sentence, and presumably also by judicially developed manuals, these instruments must have been significant drivers of this change. This should temper the criticisms of those who continue to oppose these instruments. In any event, one must doubt whether their opposition is justified.

In essence, the critics claim that guidelines and manuals breed a schematic approach to sentencing in which the great variety of sentencing factors of relevance in individual cases is impermissibly reduced to just a few,[73] with the consequence that the sentencer fails to live up to the legislative requirement of an individualised sentence. What the guidelines approach in particular neglects is the impression the judge forms of the person of the offender which, it is asserted, should be central to the sentencing decision.[74] These criticisms sometimes culminate in the claim that guidelines- or manual-based sentencing is undignified and intolerable.[75]

Such criticisms mix valid and invalid arguments. As to the latter, the impression conveyed in these criticisms that sentencing which relies on the relatively few criteria picked out in these instruments diminishes sentencing considerations that judges would otherwise be engaging in in these cases, is almost certainly false. Criminological studies of sentencing of serious crimes of interpersonal violence, for

[71] Strongly opposed to guidelines are e.g. Jagusch, "Strafzumessungsempfehlungen von Richtern im Bereich der Straßenverkehrsgefährdung" (1970) 23 *Neue Juristische Wochenschrift* 401 ff.; Jagusch, "Gegen Strafzumessungskartelle im Bereich des Straßenverkehrsrechts" (1970) 23 *Neue Juristische Wochenschrift* 1865 ff. For the views of judges, see Streng, fn. 8 above, p. 102 f.

[72] See e.g. Maurer, fn. 16 above, pp. 43 ff., 62 f.

[73] Jagusch, "Strafzumessungsempfehlungen", fn. 71 above, p. 402 f. See further Peters, "Grenzen des strafrichterlichen Ermessens", in *Verhandlungen des 41. Deutschen Juristentages*, vol 1 (1955), p. 34; Streng, fn. 8 above, p. 64 f.

[74] Jagusch, "Strafzumessungsempfehlungen", fn. 71 above, p. 403; see further Streng, fn. 8 above, p.68 f.

[75] Note the replies given to Streng when he questioned judges about this issue, see Streng, fn. 8 above, p. 102 f.

which no such guidelines exist, have shown that even in these cases, only relatively few factors really affect the sentencing decision – ordinarily only three, according to H.J. Albrecht who conducted the most extensive study of this kind.[76] Compared to real-life sentencing practice, guidance based on a small set of factors is thus unlikely to curtail the issues taken into account – provided, of course, that the guidance refers to the very factors that irrespective of the existence of such guidance are in practice considered the relevant ones. Judged against existing practice one could at most object to sentencing based on such guidelines that the guidelines might undesirably reaffirm and cement an unsatisfactory existing practice.

But even this objection is weakened if the guidelines are designed appropriately, in particular, if the factors are not correlated with any fixed penal values but with merely with narrowed ranges of sentencing outcome within the broad range provided for by statutory minima and maxima. Within the narrowed ranges, which guarantee a measure of uniformity, other relevant factors can still be given appropriate weight – including the impression which the judge gained of the personality of the accused, to the extent that this impression matters for his decision.

One should in any event not accord too much weight to the personal impression made by the accused.[77] There are a number of important questions for which the impression made by the accused is completely irrelevant – such as whether maintaining respect for the law requires a sentence of imprisonment or rules out the suspension of such a sentence. And even with regard to the severity of punishment it must be borne in mind that the offender is punished on account of the crime he committed beforehand and for which he is now on trial, and not on account of any impression he made during his trial. The personal dimension of the crime which is relevant to sentencing is the psychological state in which the offender was when he committed the crime. The personal impression he gives during his trial is, at best, a limited indicator of this state.[78] The impression of

[76] See e.g. Albrecht, "Die Geldstrafe als Mittel moderner Kriminalpolitik" in Jescheck and Kaiser (eds.), *Die Vergleichung als Methode der Strafrechtswissenschaft und der Kriminologie* (Berlin: Duncker & Humblot, 1980) 235, 244 ff.; see further Streng, fn. 8 above, p. 65 f.

[77] Of the same view: Streng, fn. 8 above, p. 68 f.; see also Frisch, fn. 44 above, p. 275 ff. with further references.

[78] See Bruns, fn. 5 above, p. 707 ff.; Frisch, fn. 44 above, p. 275 ff. with further references.

his personality which the offender gives at this trial, and which cannot be adequately incorporated into sentencing guidelines, matters more for questions of individual prevention, such as whether the accused can be expected not to reoffend if a sentence of imprisonment is suspended. In this regard, guidelines must leave judges sufficient room to arrive at appropriate decisions in the cases before them.

Provided guidelines comply with these requirements, they have positive effects on sentencing. They bring about some measure of uniformity while maintaining the advantages of appropriate individualisation of sentences, and do not make sentencing schematic. The real problems of guidelines lie elsewhere. First, it is hardly feasible in the near future to roll out guidelines across the criminal code for each offence. Secondly and more importantly, with respect to some offences and types of criminality, guidelines which attach definitive penal values to certain forms of commission are in view of the nature of these crimes unfeasible. This is for instance true for crimes of homicide.

If we want to increase sentencing consistency in this area, we therefore need other strategies. These strategies must ensure that the absence of objective guidelines does not lead the judge to have immediate unfiltered recourse to what he "believes to be the right sentence". After all, reliance on such unfiltered individual conviction was the root cause for the observed wide discrepancies in sentencing. To prevent immediate recourse to personal conviction, sentencers must be given some assistance to enable them to contribute to greater consistency in sentencing such matters.

4.4 *Data Collections Which Provide Sentencers with Information Regarding the Disposal of Other Cases*

As things stand, such help can only come from access to information about how other courts have decided similar cases.[79] Recommendations are unlikely to be possible given the potentially wide variation between sentences. Even the provision of information about decided cases is resource-intensive.

4.4.1 *Sentencing Databases*

A database which provides relevant information to all those interested in actual sentencing practice would have been impossible to

[79] Streng, fn. 8 above, p. 304 ff.; Streng, fn. 25 above, marginal note 768 f.; Maurer, fn. 16 above, pp. 91 ff., 219 f.

provide until fairly recently. It is only the possibility of electronic storage of and access to decided cases via the internet that has made such a system possible – at least if one is willing not to be overly ambitious. Assuming that one foregoes efforts to upload older cases and merely focuses on new judgments in respect of certain offences, it appears sufficient that the judge (who in any event now has an electronic version of the judgment) sends the judgment to a central collection point to upload, together with a form into which the judge has entered certain key descriptive elements of the case (similar in structure and detail to the case scenarios used in vignette studies) as well as information about the sentence imposed and the main factors taken into account. Through a keyword search, users of the system could then quickly look for other decided cases concerning the same offence and identify those usefully similar to the case before them. This would give users information about the sanctions that have been seen as appropriate in potentially similar cases, as well as an idea of the range of sanctions previously imposed for the offence. It ought also to be possible to access previous decisions in full, as this may be necessary to appreciate the degree of comparability between cases.

Japan provides an example of a jurisdiction where the creation of such a database has increased sentencing consistency.[80] In Japan, the impetus to create this database stemmed from a change in criminal procedure law which gave lay judges a key role in sentencing decisions for serious crimes of violence. The database was created to provide these lay judges with information about how other courts had decided similar cases. This is why the Japanese database only contains information about sentencing for serious violent crimes, but in principle the system could be expanded to cover other types of offences as well.

A database of this sort can be expected to quickly build up sufficient information content to constitute a useful resource. The Japanese experience is instructive: despite the fact that homicide offences are comparatively infrequent crimes, it only took a few years until the database contained enough information to give judges and other interested parties a good picture of how these crimes had been sen-

[80] Nakagawa, "Die Strafzumessung in der Tatsacheninstanz" in Frisch (ed.), *Grundfragen des Strafzumessungsrechts aus deutscher und japanischer Sicht* (Tübingen: Mohr Siebeck, 2011) pp. 201, 206 f., 214 ff.; Harada, "Die Überprüfung der Strafzumessung an japanischen Appellationsgerichten" in Frisch, *ibid*, pp. 237, 245.

tenced by other courts, and it has since been considered a great help by all concerned in sentencing fresh cases.[81]

4.4.2 *Possible Objections: Considering the Contribution of Databases to Greater Sentencing Consistency*

One likely objection to be made against the creation of such a database is cost. This objection is not particularly weighty. First of all, the jurisdictions which have successfully introduced such databases show that the cost issue is manageable. Secondly, the principled commitment of our polity to treating people equally should make us consider an expense that increases uniformity in sentencing worthwhile.

A more serious objection arises from doubt that a decision database would, in fact, increase sentencing consistency. Given the wide disparities in imposed sentences one presumes to exist in respect of some crimes, which any such collection of sentencing decisions would then naturally come to reflect, how can such a documentation do anything other than simply bring the disparity which exists into the open? And how useful is such a database if the collected cases contain no direct factual matches or only very few decisions concerning a particular offence?

As a matter of fact, this kind of database is not only useful in the eventuality that it reveals relatively consistent sentencing patterns. In areas where such sentencing patterns obtain, the database will mainly reduce the risk than a sentencer might impose an outlier sentence due to sheer lack of knowledge about what other courts usually do in this sort of case. But the database is also useful in cases where it reveals a wide spread of sentences for relatively similar factual scenarios. At the very least, this will make the sentencer (or other legal user) realise that the case with which he deals falls into a highly contentious area where he should be wary of reaching a quick decision without much reflection. Moreover, against such a backdrop a judge who considers treating like cases alike an important value will endeavour to contribute to greater sentencing equality through his decision. The realisation that this is usually best achieved by steering a middle course[82] will discourage him from imposing a sentence at the ex-

[81] The information in this paragraph is based on a briefing event held at Japan's Supreme Court in September 2009 that the author attended as part of a delegation of German legal scholars.

[82] This conception of the appropriate mean is already developed by *Aristotle*, Nicomachean Ethics, book 2, chapter 5 (1106a). It was taken up and developed

tremely mild or extremely severe end of the discovered de facto spread. Rather, he will be minded to impose a sentence somewhere in the middle – at least if he can personally consider such a sentence justified. Thus even the documentation of inequality can contribute to growing equality. That this is not just speculative is, again, supported by the Japanese experience. Here, the commitment to the value of equality pushed sentencers to follow a middle course; these sentences were considered by sentencers to be the most appropriate.[83] Of course, this does not exclude the possibility of outlier sentences. These must where necessary be dealt with in other ways.

Even where the database contains only very few entries for a particular offence, the information it provides is not entirely worthless in increasing sentencing consistency. If the entry happens to concern a factually similar constellation, the sentence recorded in the database provides an anchoring point,[84] which supports the development of a uniform approach provided that the second judge considers the recorded sentence in principle an appropriate one. If this is not the case and the second judge sentences the case before him significantly differently, giving reasons for his decision, then his judgment can in turn facilitate decisions by other judges who are minded to agree with him, leading with time to a uniform line from which the first judgment in retrospect questionably deviates.

Even if the database only contains distinctly dissimilar cases for a certain offence, it can still contribute to greater sentencing consistency. The existing judgments can, for instance, illustrate the gravity scale of possible cases[85] and in this way allow the judge (provided he considers the sentences in those cases were justifiable ones) to place the case before him on an appropriate point of the emerging gravity scale in view of the specific sentencing factors present in his case.

With recently created offences, it is of course possible that the database does not yet contain any judgments applying the offence. Even so, indirect guidance may come from decisions concerning similar offences. In any event this is no argument against the con-

---

Footnote 82 continued
further in the context of sentencing by Spendel, *Zur Lehre vom Strafmaß* (Frankfurt: Klostermann, 1954) pp. 170 ff., 172 f. and Bruns, fn. 5 above, p. 85.

[83] This information is again based on the public briefing event mentioned in footnote 81 above.

[84] On the anchoring effects of existing judgments and submissions for subsequent decisions, see above fn. 67 and accompanying text.

[85] See the discussion in Part III. 3 above.

tribution made by such a database to greater uniformity in sentencing but merely shows that the existence of such a database cannot solve every problem. Some deficits will persist, and we should think about how to overcome these deficits.

4.4.3 *Optimising the Database: Persistent Limitations*

The two issues just identified – the lack of a clear line when there is a wide variation between the fact patterns and sentences of decided cases, and the absence of jurisprudence for newly created offences – could be addressed to some extent. Where variation is the issue, commentaries should be provided which summarise and systematise the relevant judgments and on this basis put forward suggestions for an appropriate, more uniform line. In relation to newly created offences, the database could contain excerpts from legislative materials concerning sentencing and any academic comment from the time when the offence first came into force that might exist.

Of course, these are merely optimisations. They do not determine the usefulness of the database. Whether the database can fulfil its potential depends crucially on the willingness of judges to use it and to contribute with their own judgments to the formation of a consistent and generally accepted approach over time. If this willingness is lacking, or if deep disagreement persists as to what is an appropriate sentence in respect of certain fact scenarios, only external control and correction through the appellate courts can provide a remedy.

V THE CONTRIBUTION OF THE APPELLATE COURTS TO GREATER UNIFORMITY IN SENTENCING

In Germany, it was for a long time the accepted view that it was not within the appellate courts' function to interfere with the severity of an imposed sentence. It is only in recent decades that changes in appellate court practice have led to the appellate courts exerting influence over sentencing severity – and particularly during the last decade insofar as the jurisprudence of the Federal Court of Appeal, the most significant criminal court of appeal in practice,[86] is concerned.

[86] German criminal procedure distinguishes between criminal cases that are tried at first instance before the lower local court (*Amtsgericht*), and before the district court (*Landgericht*). The *Amtsgericht* deals with lower-level crimes and certain mid-level crimes where the concrete punishment upon conviction is not expected to ex-

5.1 *Developments in the Jurisprudence of the Appellate Courts Regarding the Measure of Punishment*

The legislative regime concerning the higher criminal courts of appeal empowers these courts to set aside judgments based on an error of law. Such error can arise both through misapplication and through non-application of an applicable provision. Whether the trial court's findings of fact are correct is, by contrast, not a question within the remit of the higher appellate courts; appeals to these courts cannot be based on any assertion of incorrect findings of fact but only on an assertion of incorrect application of law.[87] This perspective, when applied to questions of sentencing, initially generated the following conclusions: A court competent only to hear appeals on points of law could set aside a sentence imposed by the trial court only if that sentence was not open to the trial court to impose – for instance, it imposed a sanction in excess of the maximum permissible punishment for the crime committed, or in some other way violated a legal provision. Provided this was not the case, the imposition of the sentence was considered not to raise any questions of law, which turned sentencing into the province of the trial judge.[88] The higher appellate courts would not hear submissions by the appellants (be they convicted accused or dissatisfied prosecutors) that the punishment imposed in the case under appeal, while not outside the applicable legal minima and maxima for the offence(s) under sentence, was nevertheless too high or too low. This position did not change even after it had been accepted that a sentencing decision that overlooked relevant, or invoked irrelevant, sentencing factors could involve a violation of law. For the first sixty years or so after the coming into force of the federal Code of Criminal Procedure (*StPO*) in 1879, the sen-

---

Footnote 86 continued
ceed four years of imprisonment. Most very serious offences, as well as medium-to-serious cases where the expected punishment exceeds four years, are tried before the *Landgericht*. From the *Amtsgericht*, an appeal which entitles the accused to a full rehearing of the facts (a type of appeal called *Berufung*) to the *Landgericht* is possible, and so is a (further or immediate) appeal, on points of law only, to the Regional Courts of Appeal (*Oberlandesgericht*). In the more serious cases which are tried at the *Landgericht* at first instance, only an appeal on points of law is permitted, and this appeal goes straight to the Federal Court of Justice (*Bundesgerichtshof; short BGH*). Appeals on points of law only in criminal matters are called *Revision*.

[87] Compare Hahn (ed), *Die gesamten Materialien zu den Reichjustizgesetzen*, vol. 3, 2nd edn (Berlin: v. Decker, 1885) p. 249 f.

[88] Hahn, fn. 87 above, p. 250 f.

tencing decision remained effectively untouchable by appellate courts hearing appeals on a point of law.[89]

A first change occured soon after the end of the second world war, when Germany was under allied control and Allied Control Council Law No. 1 prohibited the imposition of cruel, disproportionately severe or unjust punishments.[90] Consequently, sentences were now checked on appeal for compliance with this requirement.[91] Subsequently, academic theorisation of sentencing as constituting not a factual determination but rather a question of law[92] promoted further developments in the jurisprudence of the appellate courts.[93] While the Federal Court of Justice (*BGH*) continues to stress that the determination of the sentence is the preserve of the trial judge and can only be reviewed by the higher appellate courts for its compliance with the law,[94] the court's more recent case law recognises that an inappropriately severe or mild sentence can as such constitute an error of law.[95] For instance, in a judgment of 1976 the court found an error of law when the perpetrator of a crime that fell within the range of frequently encountered ordinary cases for the relevant crime had been sentenced to a punishment close to the mid-point of the legal sentencing range for this offence.[96]

Later the *BGH* developed in particular two lines of argument which enabled it to set aside judgments in cases where the sentence imposed by the trial court appeared problematic even though no other errors of law were apparent. The first line of argument takes up the idea that in certain types of cases there is an observable established sentencing pattern which the appellate court is aware of and approves of. If this is the situation, the appellate courts will set aside sentences which, with no or insufficient explanation, veer away from

[89] See further Frisch, fn. 43 above, p. 259 f.

[90] See Allied Control Council Law No. 1, Article 4 no. 8.

[91] See e.g. OGHSt 1, 172, 174; further references in Frisch, fn. 43 above, p. 267.

[92] See further Frisch, fn. 43 above, p. 262; Frisch, fn. 44 above, pp. 114 ff., 179 ff.; Frisch, "§ 337 StPO" in Wolter (ed.), *Systematischer Kommentar zur Strafprozessordnung*, 4th edn. (Cologne: Heymanns, 2014), marginal notes 147 ff., 168 ff.

[93] See further, Frisch, fn. 43 above, p. 267 ff. and Frisch, fn. 92 above, marginal notes 158, 174–176.

[94] See e.g. BGHSt 17, 35, 36; BGHSt 29, 319, 320; BGHSt 57, 123, 137 and (with further references) Frisch, fn. 92 above, marginal note 148.

[95] So already Bruns, fn. 5 above, pp. 82, 714.

[96] BGHSt 27, 2, 4 f.; see also BGH NStZ-RR 2003, 52, 53; BGH StV 2010, 418 and (with further references) Frisch, fn. 92 above, marginal note 174 ff.

the established pattern.[97] The error of law in these cases lies in the insufficiency of the reasons given for the deviation from established practice. Similarly, the *BGH* will set aside a sentence when the trial court fails to provide a sufficient explanation for why an appellant who, as a co-accused who played a role in the commission of the offence comparable to another accused, was given a markedly more severe punishment than the other accused.[98] A different argument is given in cases where, due to the great variability between the cases or for other reasons, no observable sentencing pattern has (yet) evolved. Here, appellate courts considering the sentence imposed by the trial judge inappropriate resort to finding that the punishment exceeds (in either direction) the limits of what can be justified in the particular case. In effect, this means that the appellate courts now consider the imposition of an unjustifiable (overly severe or lenient) punishment, as such, to constitute a reversible error of law.[99]

Recent studies of sentencing practice show that the greater willingness of the appellate courts to set aside sentences on appeal has affected sentencing practice – ensuring that the trial courts remain within certain zones of severity in comparable cases, owing to the realisation that sentences outside these zones are likely to be set aside on appeal.[100] This intensified appellate control of the appropriateness of sentencing outcomes, together with the standardising effects of prosecutorial directives and judicial manuals for certain extremely frequently committed offences, are the most plausible reasons why recent criminological studies of sentencing no longer reveal large sentence disparities comparable to those observed in studies conducted in the 1950s and 1960s.[101]

The question that still needs to be asked is whether the appellate courts could do more to increase sentencing equality. My view is that the influence of the appellate courts could be increased in two ways.

---

[97] See further Frisch, fn. 92 above, marginal note 175 and Streng, fn. 25 above, marginal note 662.

[98] See the references quoted in fn. 25 above.

[99] See e.g. BGHSt 45, 312, 318 f.; BGHSt 53, 71, 86; BGHSt 57, 123, 130 f. and (with further references) Frisch, fn. 92 above, marginal notes 174–176.

[100] For a detailed analysis of the case law, see Maurer, fn. 16 above, pp. 128 ff., 140 ff., 147 ff.

[101] Recall also the uniformity-promoting effects of prosecutorial guidelines addressed in Part IV.3. above.

5.2 *Optimising the Role of the Appellate Courts in Ensuring Sentencing Equality*

One possible way in which appellate courts could optimise their efforts to increase sentencing equality relates to the identification of the reference point of an established sentencing pattern that has developed in comparable cases. It has repeatedly been questioned where the appellate courts derive their knowledge of the usual punishment from and how well-founded their notion is of the usual punishment for certain types of crimes. Clearly, the appellate courts' knowledge of the sentences given by trial courts is necessarily somewhat coincidental and limited. This limits the confidence with which particular sentences can be set aside on the ground that the sentencing judge has veered too far afield from what is usual in this sort of case, without providing sufficiently strong reasons for this. Such reasoning by the appellate courts would be much strengthened if a sentencing judgment database of the kind I described above were to be created. Such a database would not only assist trial judges in identifying an appropriate sentence but would help the appellate courts to form a reliable impression of the sentences given in the lower courts for certain types of cases. Unreasonable departures from what other courts do in this sort of case can then be identified by appellate courts long before reliable quantitative data has been collected or before a settled sentencing practice has emerged.

The identification of the width of the range of sentences given for certain types of offences which a sentencing database would make possible, is again an advantage of such data collection that benefits not only trial courts. It was argued above that, for the trial judge, the discovery of large discrepancies between existing sentences provides an impetus to think of his own judgment as contributing to the development of a generally acceptable line. The same impulse will no doubt be felt by appellate judges concerned to increase sentencing equality as part of their overall task to ensure equality before the law.[102] If the appellate court is of the view that the trial judge has, in the case before it, imposed a sentence capable of indicating a reasonable approach to be followed by other courts in the future, it could say so clearly with a view to encouraging other judges in future cases to follow this line. On the other hand, an appellate

[102] On the function of appellate courts to ensure uniformity of application of the law, see further Frisch, "Vor § 333 StPO" in Wolter (ed.), *Systematischer Kommentar zur Strafprozessordnung*, fn. 92 above, marginal note 14 ff. with further references.

court which is of the view that the sentence imposed in the case before it is problematically close to the most or least severe sentences registered in the data collection, or indicative of an idiosyncratic approach of the trial judge, and thus unlikely to be considered acceptable by other judges called upon to decide future cases, should highlight this and set the judgment aside. At present, this is done as a matter of course in cases where the trial court's judgment is marred by other errors of law (say, a misappreciation of relevant sentencing factors) which affect the sentencing decision. The appellate courts should, however, also be encouraged to set aside – in the interest of contributing to the development of an appropriate general line – a sentence imposed by a trial court, merely because they consider the punishment imposed in the particular case to be so close to unjustifiable that it would be inappropriate for the judgment to exert any persuasive pull on other judges in future cases. Arguably, the appellate courts have the authority to do so in that the trial court's sentencing judgment lacks an important quality of decisions made in accordance with the law – namely, that they be capable of generalisation into a general rule for the disposition of comparable cases.[103]

Of course, it could be objected that the appellate courts would then arrogate to themselves questions and decisions properly left within the province of the trial courts. But this objection is weak. It is part of the function of appellate courts to ensure equality before the law. There is no reason whatsoever why a panel of three or five experienced appellate judges should not be able to appreciate, and be called upon to consider, whether a given sentence is appropriate in a particular case – something that we unhesitatingly consider every trial judge to be able, and be called upon, to do.

Against this background we should welcome the fact that the appellate courts have recently been prepared, not just to watch over the appropriateness of sentences in individual cases and set aside sentencing decisions if need be, but also to state clearly what they would consider an appropriate sentence to be and to formulate guidance for the adequate assessment of certain types of cases in the future. This path has been charted in an exemplary fashion in a series

[103] On the possibility of universalisation of a rule as a criterion for its legal nature, see Kant, *Metaphysik der Sitten, Rechtslehre, Einleitung in die Rechtslehre*, Werkausgabe, vol. 8 (Weischedel ed) (Frankfurt: Suhrkamp, 2012) § C; see further Henkel, *Einführung in die Rechtsphilosophie* (Munich: Beck, 1964) p. 354 ff.; Radbruch, *Rechtsphilosophie*, 6th edn (Stuttgart: Koehler, 1963) pp. 127, 128, 206.

of *BGH* judgments on sentencing for tax evasion and related offences, where unusually detailed, systematic and comprehensive guidance was formulated on how such cases should be sentenced.[104] This guidance not only clarifies when a case is no longer appropriately dealt with by a fine but has crossed the custody threshold,[105] but also addresses the appropriate length of custodial sentences to be imposed in light of factors such as the sums involved and the period of time over which the offence has been committed.[106] In this context the appellate court has also provided specific guidance, going well beyond the general principle stated in the applicable legislative provision, on when suspension of a sentence of imprisonment is still appropriate.[107] This recent jurisprudence has made an important contribution to greater equality of sentencing in these matters. It is to be hoped that these efforts of the higher courts will not remain restricted to sentencing for tax offences but will be extended to other areas of criminal law as well.

[104] See particularly BGHSt 53, 71, 84 f.; BGHSt 57, 123, 130 f.; BGH NJW 2012, 1015, 1016; BGH StV 2012, 219; for details see Rastätter, fn. 65 above, pp. 143 ff., 151 ff.

[105] See BGHSt 53, 71, 86 f.

[106] See BGHSt 53, 71, 86; BGHSt 57, 123, 131 ff.

[107] See BGHSt 53, 71, 86; BGHSt 57, 123, 131 f.