CONCLUSIONS
The main aim of this study was to empirically analyze consistency of international sentencing. Focusing on the two most active international criminal tribunals nowadays – the International Criminal Tribunal for Rwanda (ICTR) and the International Criminal Tribunal for the Former Yugoslavia (ICTY) – it empirically described the current international practice of sentencing for war crimes, crimes against humanity and genocide and analyzed its consistency. The three basic dimensions of consistency of international sentencing practice – consistency in approach, consistency in outcome and systemic/cross-institutional consistency – were discussed. The ICTY and ICTR case-law was analyzed and the data on all defendants convicted and sentenced by the ICTY or ICTR up to April 2011 were collected. This concluding chapter provides a summary of the main results, discussion of their implications and methodological limitations of the study followed by a discussion on the need of drafting international sentencing guidelines. The chapter concludes by suggestions for future research on international sentencing.

1. **Summary of the Results**

The first part of this thesis focused on consistency in approach of international sentencing. Chapter I analyzed the ICTY and ICTR judges’ sentencing narrative and argumentation. Since the positive law provides only very limited guidance with respect to sentencing, the analysis focused on sentencing argumentation of the ICTY and ICTR judges in individual cases. The case law of both Tribunals was examined to identify consistent patterns and any disparities in judges’ sentencing arguments across cases. The cases decided by the ICTY and ICTR judges up to April 2011 were included in the analysis (N=124). The structure of sentencing reasoning and general principles of sentence determination seem to be mutually influencing under both jurisdictions. Judges from one tribunal often refer to the case-law of the other in their general sentencing considerations, thus developing a common ICTY-ICTR narrative. The Tribunals share a joint Appeals Chamber which also arguably contributes to the development of a ‘common’ jurisprudence. On the one hand, it turned out that on a general level, a set of sentencing principles is consistently discussed and emphasised by the ICTY and ICTR judges: the most important being the primacy of gravity of crimes in sentence determination and the principles connected to the gravity evaluation: proportionality, gradation and totality. The principle of individualization complements this general framework. Consequently, sentence severity at the ICTY and ICTR seems to be primarily determined by factors relating to the
gravity of the crime (by applying the principles of proportionality, totality and gradation and assessing aggravating factors), and the sentence is then adjusted by taking into account the individual circumstances of the offender, i.e. mitigating factors (‘principle of individualization’). On the other hand, however, the analysis detected some instances of disparities across cases. Inconsistencies in the ICTY and ICTR judges’ reasoning were identified especially with respect to the detailed application of these general principles to individual facts of each case. In particular, the ICTY and ICTR sentencing case law largely varies with respect to a selection of factors relevant for the gravity assessment and distinction between the notions of ‘gravity of crimes’ and ‘aggravating factors’. Second, differences between individual cases exist as to whether a particular mitigating/aggravating factor aggravates/mitigates a sentence and its significance for meting out a sentence in a particular case. 

The analysis conducted in Chapter I served as a basis for building up an empirical model of international sentencing. The ICTY and ICTR judges do not provide a detailed breakdown of how a sentence was calculated in individual cases and how the individual factors influenced sentence severity. It is, therefore, not clear from the case law how the general principles and individual sentencing factors translate empirically into sentence determination. The empirical analysis of sentencing conducted in the rest of this thesis, therefore, shed additional light on the sentencing of international crimes and revealed practical implications for sentence severity of the general sentencing principles consistently emphasised in the ICTY and ICTR case law. It turned out that also empirically there have emerged consistent patterns in the sentencing outcomes at the Tribunals and a set of legal factors is used in a consistent and predictable way at the ICTY and ICTR.

The second part of this thesis, therefore, focused on consistency in outcome of international sentencing and empirically analyzed the sentencing practice of the ICTY and ICTR, respectively. Chapter II dealt with the sentencing outcomes at the ICTY. It empirically tested to what extent legal factors stemming from the general principles of sentence determination emphasized in the ICTY case law predict sentence length in individual cases. The set of general principles of sentence determination consistently emphasized by the ICTY and ICTR judges had been identified in Chapter I - the primacy of the gravity of crime (and related principles of totality, proportionality and gradation) and sentence individualization. The selection of independent variables included in the statistical analysis in Chapter II was inspired by these
main principles. The gravity of crime was operationalized by variables measuring (i) a category of crime of which a defendant was convicted, (ii) number of guilty counts, (iii) number of aggravating factors, (iv) modes of individual liability and (v) rank in the military or political hierarchy. The individual circumstances of the offender were represented by the number of mitigating factors judges accepted in individual cases. Guilty plea was also added to the model as a separate variable. Data were collected on all defendants sentenced by the ICTY up to August 2008 (N=63) and analyzed using a multiple regression analysis. The results demonstrated that well-interpretable and legally relevant patterns in the ICTY sentencing practice have emerged: Those convicted of more extensive criminal activities are punished more severely than perpetrators of isolated, single acts. The analysis also suggested a distinction in sentencing practice between crimes against humanity and war crimes. Those guilty of crimes against humanity receive longer sentences. The high ranking defendants are sentenced by the ICTY to substantially longer prison terms than the ordinary low ranking offenders. Furthermore, the ICTY judges reduce sentence if they accept factors in mitigation and finally, those who instigate others to commit crimes are punished more severely than all the other participants in the atrocities. This study has demonstrated that despite the loose legal regulation, legally relevant patterns in the ICTY sentencing jurisprudence have emerged – on the basis of the combination of five legal factors 60% of sentence variation can be explained.

In Chapter III attention turned to the ICTR and its sentencing practice. In contrast to the ICTY, the ICTR has dealt primarily with genocide charges and indeed, almost all ICTR defendants were convicted of genocide and many of them sentenced to life imprisonment. The basic principles of sentence determination (already discussed in the Chapter I) at the ICTR were first reviewed and then, ICTR sentencing practice was empirically analysed with a focus on the relationship between sentence severity and gravity of crime. Factors relevant to the assessment of crime gravity such as the extent of crimes, the category and type of convicted crime, the way and degree of defendants’ participation in crimes and their enthusiasm in committing crimes were included in the analysis. Data were collected on all defendants sentenced by the ICTR up to March 2011 (N=44). Due to the high prevalence of life sentences at the ICTR and the low number of defendants sentenced by the ICTR so far, it was not feasible to use multiple regression analysis or any similar statistical method. The ICTR data were thus investigated using a multiple correspondence/homogeneity analysis (HOMALS) – a non-linear, multivariate
technique used to explore and summarize categorical data by searching for multivariate associations between variables. First, the HOMALS solution distinguished between two categories of cases on the basis of a defendant’s active or passive involvement in crime: the passive participants are subjected to generally short sentences while more severe sentences are reserved for more active contributors. Second, the sentence severity of the active contributors is then determined by a degree of culpability which in turn depends on three main considerations: (i) the defendant’s position in the conflict; (ii) the extent of involvement in atrocities (whether a defendant merely facilitated crime committed by others, or personally committed or organized the violence), and (iii) overall harm (whether a defendant is found guilty of participating in a limited number of or multiple/recurrent attacks). On the one hand, the most severe sentences at the ICTR were handed out to defendants who occupied the most senior positions in the civil or military hierarchy and organized massacres from behind their desks or to authority figures who participated in crime with particular zeal and sadism. No low-ranking defendant, no matter how enthusiastic or zealous, has yet been sentenced to life imprisonment by the ICTR. Consequently, leaders seem to be regarded as the most culpable by the ICTR judges. On the other hand, the high rank and influential position of a defendant is not the only consideration in sentence determination at the ICTR. ICTR judges seem to weigh the position of defendants in the overall conflict against the level of their involvement in crime. If their crimes are limited in scope and/or degree (if, for example, the crime basis is limited to a single attack or the degree of involvement is very limited), sentences tend to be more lenient. Therefore, the results demonstrated that patterns in ICTR sentencing could be identified. The empirical analysis indicated that the principles discussed in the ICTR sentencing jurisprudence are generally reflected in the ICTR sentencing practice and ICTR sentences seem to be gradated according to the increasing seriousness of committed crimes and a defendant’s culpability.

The final part of the thesis focused on cross-institutional/systemic consistency of international sentencing and compared the sentencing practice of both Tribunals. Chapter IV presented a comparative, descriptive analysis of the ICTY and ICTR sentences and offered a detailed examination of the differences in the sentencing practices of the two Tribunals. The sentences handed down by the ICTY and ICTR up to June 2010 were included in the analysis (N=111). Median sentences were computed for individual categories of cases depending on the category
of crime, type of underlying offence, scale of crime, mode of individual liability, rank of a defendant (and his role in crime). The differences in sentence severity among individual groups of cases and between the Tribunals were then compared and discussed. Furthermore, the most frequent aggravating and mitigating factors cited by the ICTY and ICTR judges were identified and compared. The analysis built upon the already well-known fact that there are large differences in sentence severity between the Tribunals - the ICTR sentences are substantially longer compared to those issued by the ICTY - and tried to identify main reasons for this divergence. It was demonstrated that the different composition of cases at the ICTY and ICTR, respectively, could be perceived as the main reason for these differences in sentence severity.

First, the analysis suggested that there are indicia of an empirical ordering of international crimes in terms of sentence length with genocide at the top, followed by crimes against humanity and war crimes at the tail. The dominance of genocide convictions at the ICTR was identified as one of the primary reasons for the more severe ICTR sentences. In contrast, the majority of the ICTY defendants are convicted and sentenced for crimes against humanity and/or war crimes and consequently sentenced to comparably more lenient sentences. Second, similarly to Chapter III, the analysis revealed that at both Tribunals sentences seem to be graded according to the scope of the crime, the role of a defendant and his/her rank. The results indicated that those responsible for regional or national campaigns of persecution, extending over longer periods of time and involving multiple victims, were subjected to the severest sentences. In general, the high ranking figures organizing crime from above were sentenced to the lengthiest sentences at both Tribunals. Since the majority of defendants that stood trial at the ICTR were orchestrators and organizers of slaughter and violence (members of government and other high ranking figures), this fact could also explain the more severe sentences at the ICTR. In contrast, the ICTY has dealt with a comparatively higher number of low ranking hands-on executioners of persecutory campaigns. Another factor that could account for the difference in sentence severity identified by the analysis is the relatively limited range of crimes prosecuted by the ICTR. The crime base and composition of convictions is much less varied at the ICTR – all accused are convicted for killing and/or serious violence. These are arguably the most serious violations of international criminal law. At the ICTY, convictions include also generally less serious offences, such as crimes against property. Finally, the overview of the most common aggravating and mitigating factors revealed that there do not
seem to be any substantial differences between the Tribunals. With few exceptions, similar mitigating/aggravating circumstances are accepted by judges at both jurisdictions.

In the final Chapter V the consistency of outcome of the sentencing at both Tribunals was examined. Since the descriptive analysis presented in Chapter IV could not fully uncover the impact of the interrelated sentencing determinants on sentence severity as this is based on the interplay of a multiplicity of sentencing factors, the final chapter focused on a multivariate analysis of ICTY and ICTR sentencing. Next, the analysis specifically addressed cross-institutional consistency of international sentencing and examined possible differences in the influence of sentencing factors on sentence length between the Tribunals. The sentencing model tested was similar to the model employed in Chapter II. The data were collected on all defendants sentenced by the ICTY and ICTR up to June 1 2010 (N=111). The collected data were analysed using multiple regression analysis. Given the N for the ICTR, it was not possible to run separate models for the two Tribunals and the data had to be combined in one dataset. The possible differences in sentence mechanisms were analyzed (i) bivariately, by testing for the differences across the Tribunals in associations between sentence length and independent variables and (ii) multivariately, by including interaction terms in the conducted regression analysis. The results indicated that the sentencing practice of the ad-hoc Tribunals appears to be rather consistent with no major differences between the ICTY and ICTR regarding the influence of individual sentencing factors on sentence severity. The set of statistically significant predictors was very similar to the model tested at the ICTY sentences. Empirically consistent (and legally relevant) patterns have emerged in the sentencing practice of both Tribunals. Six of all the tested legal factors statistically predict sentence length at the ICTY and ICTR: Those convicted of genocide have been subjected to the severest sentences. Compared to shorter sentences handed down to defendants convicted of crimes against humanity, genocide seems indeed to be treated as ‘the crime of crimes’ for the purposes of sentencing. The rank of a defendant stood out as another important predictor of sentence length. The high ranking defendants were subjected to the lengthiest sentences compared to their lower ranking followers. More extensive criminal activity also leads to longer sentences. Those convicted of multiple crimes are punished more harshly than those convicted of isolated, single acts. The results showed that ‘on average’ each additional guilty count adds to sentence severity. The number of counts stood out as the only predictor with an arguably different effect on sentence length across the Tribunals: a guilty count at the ICTR results in a larger increase in sentence length compared
to the ICTY. This difference could be related to the fact that separate counts reflect more serious and/or extensive criminal conduct at the ICTR: in contrast to the ICTY all the ICTR defendants were convicted of killings and violence, as demonstrated in Chapter IV. Finally, the analysis revealed consistent patterns of (i) a sentence extension every time judges accept an aggravating factor and (ii) a sentence reduction when mitigating factors are identified. The model accounted for 70% of the variation in sentence length at the ICTY and ICTR practice.

2. Discussion of the Results

This thesis has demonstrated that despite all the criticism raised against the sentencing of the ad hoc international criminal tribunals, consistent and predictable patterns have emerged in their sentencing practice. Taking into account the results of all the conducted studies the following sentencing patterns have been discovered in the ICTY and ICTR sentencing practice: (i) an empirical ordering/hierarchy among different categories of international crimes with genocide being considered the most serious category of crime, followed by crimes against humanity and war crimes; (ii) a culpability level of a perpetrator is assessed by the ICTY and ICTR judges taking into account primarily a defendant’s position in the overall state hierarchy, his role and the scope of his/her crime with high ranking organizers convicted on multiple crimes being subjected to the most severe sentences; and (iii) relevance of individualization of sentences with aggravating and mitigating factors cited by the ICTY and ICTR judges accounting for an increase/reduction in sentence length.

2.1. Hierarchy of International Crimes

According to the empirical results it seems that genocide is indeed considered ‘the crime of crimes’ for the purposes of international sentencing and those convicted of genocide are sentenced to the harshest sentences. Conviction for crimes against humanity entails more lenient sentences and war crimes seem to be punished the least. Generally, on the one hand, academics have theoretically argued for the hierarchy of individual categories of international crimes based on the analysis of their gravity in abstracto with genocide at the top of this severity scale; followed by crimes against humanity; and war crimes as the least serious violations of
international criminal law. Our analyses indicated that also in an empirical sense this distinction holds true. On the other hand, the ICTY and ICTR judges, however, dismissed the idea of the crime hierarchy and argued that all the categories of international crimes are equally serious. The judges instead emphasised the gravity of crimes in concreto, i.e. degree of harm actually caused by a defendant’s criminal conduct. Therefore, according to the judges, the legal classification of an act as genocide does not matter so much. In practice, many defendants convicted of genocide were the organizers of a campaign of genocide and usually convictions of genocide encompassed brutal treatment and killings of a multiplicity of victims. Conversely, the defendants convicted exclusively of war crimes tend to be lower ranking individuals convicted on the basis of criminal activities never amounting to the organization of a state/regional persecutory campaign – in many cases these defendants were convicted as superiors for their omissions or for only ancillary activities. Therefore, it appears that genocide convictions are indeed in terms of the gravity in concreto the worst cases with the greatest amount of harm and this is what brings about the severest sentences. Convictions for war crimes, in general, encompass comparably less serious cases entailing the least amount of harm and thus the most lenient sentences.

In order to assess the relationship between sentence severity, legal classification of a criminal conduct and its particular circumstances, it would be necessary to conduct a more detailed comparative analysis incorporating the particularities of each case, such as type of underlying offence (e.g. murder, violence, or property crimes), the number of victims and some other quantifications of harm caused by an offender. The extent of criminal activity/the scope of crimes and different types of underlying offences were examined descriptively in the Chapters III and IV of this manuscript. Indeed, at the ICTR where most of the genocide convictions are entered all defendants are convicted of killing and violence and the majority is found guilty for recurrent and extensive crime. However, to investigate a link between the gravity in concreto and sentence severity, a more detailed multivariate statistical analysis of sentencing incorporating not only the legal classification of an act in a sense of the category of crime but also the variables such as underlying offences, extent of harm or number of victims would be necessary. Arguably, such an analysis would be difficult to conduct for three main reason (i) the

limited number of sentences (in a statistical sense) handed down by the Tribunals so far substantially limits possibilities of any more complicated and detailed modeling; (ii) as discussed below, it is extremely difficult to come up with a measure quantifying harm of a defendant’s criminal conduct in case of international crimes being the crimes of international concern attacking ‘humanity’ as such. The number of victims seems to be one possible measure of harm in cases of international crimes. However, certain types of international crimes do not necessarily entail direct victims that can be counted such as for example persecution on the basis of destruction of towns, villages or other public and private property or many war crimes such as attack against civilian objects or attack on undefended towns; and (iii) even for cases where it is possible to estimate the number of direct victims, exact numbers of victims affected by a perpetrator’s criminal activities are sometimes lacking in the judgments.

2.2. Culpability level

At the ICTY and ICTR, the culpability level of a defendant seems to be primarily determined by his/her position in the overall state hierarchy combined with the assessment of the level of his involvement and scope of his/her crimes. All the conducted studies confirmed that those at the highest leadership levels are subjected to the severest punishment at both Tribunals. In the two regression studies the rank of a defendant stood out as one of the major predictors of sentence severity at the ICTY and ICTR. The highest authority figures are subjected to a more severe punishment than their respective followers – those without any influence whatsoever in the overall conflict. This finding indicates that those orchestrating crime at the top leadership levels are indeed considered the most culpable for the purposes of sentencing. Theoretically, international crimes are perceived as a manifestation of collective, large-scale and systematic violence. According to criminological theories explaining social dynamics of international crimes the role of the top organizers is instrumental to the development of large scale criminal systems that commit international crimes and the top leadership figures are those most to be blamed for international crimes.² These theoretical considerations seem to be reflected in the empirical sentencing practice of both ICTY and ICTR. However, the conducted analyses also

indicated that there is no further gradual decrease in sentence severity for middle and low ranking offenders at the Tribunals sentencing practice.³ Possible explanation for this phenomenon can be that in the cases of low ranking defendants, considerations of the role of offender in the overall conflict are overshadowed by considerations of cruelty and depravity of committed crimes and zeal of their executioners. As opposed to their direct, middle ranking, superiors, the low ranking offenders are in the majority of cases the actual perpetrators committing brutal atrocities and are in light of the actual brutality and depravity of the particular attacks assigned equal or even more blame than their direct superiors by the judges. Consequently, the exact role of a defendant in crime, his/her enthusiasm and zeal in their execution also influence the assessment of his/her culpability.

The role of a defendant in crime is also expressed by modes of individual liability according to which he/she is convicted. The analyses, however, did not reveal any conclusive empirical patterns in the ICTY and ICTR sentencing practice in this respect. It seems that the ICTY delivers more differentiation in sentence length depending on a mode of individual liability as demonstrated by the descriptive analysis and the results of the regression analysis on the ICTY data where the analysis revealed statistically significant or emerging patterns with respect to sentencing instigators, hands-on perpetrators and superiors. However, similar patterns were not found in the multivariate analysis of the sentences of both Tribunals. Several reasons could explain this: (i) as discussed below, the limited number of cases decided by the Tribunals so far and included in the analysis might, in theory, have caused instability in the models. However, empirically, despite the limited number of cases compared to the number of predictors included in the analysis, the presented models are remarkably stable and robust. The truth is, as discussed in Chapter I, that the ICTY and ICTR case law on relationship of individual modes of liability and their impact on sentence severity is not very well developed either. This theoretical confusion could have also caused the lack of empirical patterns in the ICTY and ICTR sentencing practice in this respect. Furthermore, with respect to some modes of liability, only a

³ Based only on results of descriptive analyses, this phenomenon is identified only at the ICTY. At the ICTR, it would appear that there is a clear gradation of sentence severity according to the rank of a defendant. However, results of the regression analysis conducted on the combined data of both Tribunals also did not confirm statistically significant differences between sentence of a low and middle ranking defendants (middle rank as a predictor just missed the statistical significance). However, since the dataset was not balanced – the ICTY cases were more numerous – it is possible that the estimate was influenced more by the ICTY data. For similar results for the ICTY see also U. Ewald: ‘Predictably Irrational’ - International sentencing and its Discourse against the Backdrop of Preliminary Empirical Findings on ICTY Sentencing Practices, 10 International Criminal Law Review (2010) Number 3, 365-402.
handful of defendants have been convicted on the basis of these liability theories and it is thus difficult to statistically examine differences in sentence length for these cases occurring scarcely in the datasets; (ii) determinations with respect to individual modes of liability are made with regard to each individual count and it is often the case that defendants are convicted on the basis of multiple liability modes. However, the analyses presented in this manuscript were conducted on a case level (not distinguishing among individual counts) and only reflected the fact whether a defendant was convicted of a particular mode of liability coded as a dichotomous variables (YES/NO). In practice, many defendants are convicted for multiple counts each entailing (a) different mode(s) of liability. Therefore, on a case level there have been a large variety of combinations of liability modes and overlap among variables measuring individual liability modes in the data, which could have led to statistically insignificant results, especially taking into account a limited sample size.

Next to the rank and role of a defendant in crime, the culpability level of the perpetrator is further determined by the extent of his crime. The two descriptive studies indicated that defendants found responsible for extensive crimes, extending over longer periods of time and involving many victims, were subjected to the severest sentences. Similarly, the regression analyses confirmed that more extensive criminal activity leads to longer sentences at the ICTY and ICTR. The number of guilty counts stood out as another important predictor of the ICTY and ICTR sentences. Those convicted of multiple crimes are punished more harshly than those convicted of isolated, single acts. The results showed that ‘on average’ each additional guilty count adds to sentence severity. In practice, it is arguably the case that a sentence increase is not uniform for every count and depends on numerous factors such as the type of underlying offence and the defendant’s role therein. In order to examine this, a more detailed analysis encompassing, for example, an influence of each underlying crime of each guilty count on sentence length would be necessary.

2.3. Individualization of Sentences

Finally, the empirical analyses revealed consistent patterns with respect to aggravating and mitigating factors. Qualitatively, there do not seem to be any substantial differences in the type of aggravating and mitigating factors cited by the ICTY and ICTR judges. Quantitatively, the multiple regression analyses confirmed that aggravating and mitigating factors identified by
judges in individual cases indeed account for sentence extension and reduction. On average, sentences seem to be (i) increased every time judges accept an aggravating factor and (ii) reduced every time judges identify a mitigating factor. However, the conducted multivariate analyses could not distinguish between different types of aggravating and mitigating factors and their significance for sentence severity. Chapter I demonstrated that a large variety of circumstances has been accepted by the ICTY and ICTR judges in aggravation and mitigation of sentences. Given the limited sample size, it was not possible to distinguish among these individual circumstances and their relative weight in sentence determination.

2.4. Consistency of International Sentencing

Could we label international sentencing as consistent? There is no simple answer to this pretty straightforward question. This thesis has analysed the three main dimensions of consistency of international sentencing – consistency in approach, consistency in outcome and systemic/cross-institutional consistency.

Regarding consistency in approach, ideally, sentencing should constitute a uniform and internally coherent system where all ‘pieces of the puzzle’ fit together and point in the same direction. Clear sentencing purposes should be pronounced to guide judges in their sentence determination and set broadly defined objectives of punishment. The general sentencing principles, i.e. broad sentencing principles stating basic rules applicable across cases, should follow from the declared sentencing aims and determine what factors are relevant and indicate their importance in individual cases. Individual aggravating and mitigating factors should logically follow from the sentencing purposes and be ascribed weight in order of priority among individual sentencing aims. Furthermore, sentence determinations in individual cases should be well-reasoned, clear and transparent so that defendants could understand and appraise how their sentence was actually meted out. Chapter I demonstrated that a set of general principles of sentence determination has evolved and is consistently emphasised in the sentencing practice of the ICTY and ICTR. This is, however, a rather rudimentary basis of ‘a sentencing system’ and these principles do not limit the judges’ discretion considerably since they only set very broad constraints basically all coming down to the fact that the gravity of crimes is to be the main sentencing factor. Many aspects of sentencing reasoning at the ICTY and ICTR remain unresolved and there are discrepancies across cases. First, judges’ declarations regarding the
justifications of international punishment are not uniform and differ across cases. The ICTY and ICTR judges usually neither attempt to elaborate in detail what actually the proclaimed sentencing objectives entail in terms of sentencing in individual cases nor try to justify the relevance of individual sentencing factors with reference to sentencing objectives. Second, the concept of gravity of crime – the primary consideration in sentencing according to the ICTY and ICTR – remains unclear and confusing. There is no uniform approach in the case law as to what the gravity of crime actually entails, how it is distinguished from elements of an offence (necessary to be established for convicting a defendant) or individual aggravating and mitigating factors. Third, different benches often differ in their opinions regarding relevant aggravating and mitigating factors and their significance for sentence determination. Individual aggravating and mitigating factors are never connected to the proclaimed sentencing objectives and the ICTY and ICTR judges only very rarely indicate what weight has been ascribed to these sentencing factors.

The main problem with respect to consistency in approach, however, lies in the lack of transparency of sentencing argumentation at the ICTY and ICTR. In the majority of cases, judges pronounce one global sentence for multiple crimes with no clear indication how individual crimes and individual sentencing factors were weighted against the sentence. Nothing is known about the actual calculations of judges and how they came up with a particular quantum of sentence in individual cases. There are therefore no clear/objective benchmarks against which the exercise of the ICTY and ICTR judges’ discretion in sentencing could be assessed. Some suggestions to address this problem were offered in Chapter I.

Regarding consistency in outcome – the predictability of ICTY and ICTR sentencing, the empirical part of this thesis revealed that despite the opaqueness of sentence determinations and a certain amount of discrepancy in judges’ sentencing reasoning, a set of legal factors is actually used in a consistent and predictable way in the ICTY and ICTR sentencing practice. On their basis it is possible to predict sentencing outcomes at the ICTY and ICTR to a considerable extent (60% of sentence variation was accounted for by the ICTY sentencing model; 70% of sentence variation was explained by the model combining data of both Tribunals). Is that good enough? Could we label the Tribunals’ sentencing practice as predictable? It is very difficult to judge the adequacy of this number, both theoretically and empirically. Given the scope of the models, which are limited to examination of very general categories and include very
elementary variables, it is difficult to expect higher numbers and impossible to account for 100% of sentence variation. Possibly, a higher explanatory power would be achieved with many more cases in the datasets and with models including variables measuring more detailed distinctions among cases. Since the data forming the basis of this research actually represent the whole population of cases decided by the ICTY and ICTR at the time of writing, it was impossible to extend the datasets. Furthermore, in theory, sentencing should never be 100% predictable. If it were so, judges would be conceived of as mere automatons mechanically dispensing ‘imaginary justice’ while handing out sentences that could hardly be seen as fair. In order to be fair, sentences should not only be predictable but also individualized reflecting particularities of each case at hand: each case has its specifics. If sentencing were 100% predictable, this would mean that particular differences among offenders would go unnoticed, and such sentencing could hardly be perceived to be fair. Fairness thus demands not only predictability but a certain degree of unpredictability. Therefore, there should always remain some unaccounted variation in sentencing to allow for individualization. Empirically, our analysis can be compared to findings of similar quantitative research into sentencing consistency at a domestic level. In many countries such as the USA, concerns over unwarranted disparities in sentencing led to various reforms of sentencing law introducing e.g. determinate sentencing law or numerical sentencing guidelines. These instruments introduced a rigid sentencing regulations and limited discretionary powers of domestic judges considerably. To evaluate the success of these reforms, numerous studies evaluating sentencing predictability at a domestic level have been conducted. For example, statistical analyses conducted at the US Sentencing Commission subsequent to implementation of sentencing guidelines determined that legally relevant factors are the most important determinants of sentences pronounced by federal judges, explaining as much as 70% variation in imprisonment length. Consequently, it seems that at the international level, despite the lack of sentencing guidelines and large discretionary powers of judges, the sentences appear to be as statistically predictable as sentences in domestic legal systems with more detailed legal regulation of sentence determination. This finding could have practical implications with respect to the recent calls for necessity to limit judicial sentencing

---

discretion at the international level and for establishment of international sentencing guidelines. Are these initiatives really necessary and feasible? These questions are briefly discussed below.

With respect to systemic/cross-institutional consistency, it appeared that the sentencing argumentation seems to be very similar at the ICTY and ICTR. Judges of both Tribunals cross-reference each other’s case-law and seem to rely on similar basic sentencing principles and emphasize similar general sentencing factors. The only relevant difference lies in the fact that judges are instructed to consider the local practices in Yugoslavia and Rwanda respectively. As discussed in Chapter I, the judges have ‘dismissed’ this command and considered national sentencing practice as only one of the many (non-binding) factors in sentence determination. Consequently, the ICTY and ICTR case-law has developed fairly similarly. Also with respect to the actual sentencing outcomes, it seems that the same general sentencing factors predict sentence length at both Tribunals. Despite differences in the levels of sentence severity between the Tribunals stemming in particular from the different composition of prosecuted cases, international sentencing appears thus to be a fairly consistent system. In an empirical sense, the mechanisms behind the sentence determination at both Tribunals appear to be very similar. Consequently, consistent and legally relevant patterns have emerged in the ICTY and ICTR sentencing practice. These patterns basically follow from the general principles of sentence determination that are consistently emphasized in the ICTY and ICTR case law and a rather substantial amount of variation in sentence length at the ICTR and ICTY is accounted for on their basis. The question is whether identification of some amount of consistency in judges’ approach to sentence determination while, empirically, being able to explain 60-70% of variance on the basis of five/six legal predictors is ‘good enough’ to conclude that the sentencing practice is consistent. The answer to this question is a matter of both normative judgement and empirical realism. Several counter-arguments, however, may be presented in this respect. First, Chapter I revealed disparities among cases especially with respect to the implementation of the principle

5 To conclude that international sentencing indeed constitutes uniform and consistent system on the basis of the results of this study is, however, premature, since this thesis did not examine sentencing practice of all other internationalized but also domestic institutions sentencing perpetrators of international crimes. As opposed to the ICTY and ICTR, where particularities stemming from domestic contexts arguably play a minor role, in case of all the internationalized and domestic courts ‘the domestic element’ is arguably much stronger and the influence of a particular legal culture and domestic norms and practices could theoretically lead to fragmentation rather than unification of the international sentencing system. The results of this research could, however, be seen as the first steps on the way to assess the unity or fragmentation of international sentencing regime. In order to assess conclusively consistency of international sentencing as such it is necessary to examine sentencing practice of all these other courts and tribunals dealing with international crimes.
of individualization and treatment of aggravating and mitigating factors. Second, with respect to
certain legal factors the statistical analysis indicated discrepancies between judges reasoning and
the empirical effect of these factors on sentence length. One of the examples thereof is the
hierarchy among individual categories of international crimes. In the case law the ICTY and
ICTR judges consistently argue that there is no hierarchy among individual categories of
international crimes for the purposes of sentencing. However, the empirical analyses with no
exception suggested that empirically there are indicia of hierarchy among the categories for the
purposes of punishment with genocide at the top, followed by crimes against humanity and war
crimes.6 Third, for another set of legal factors (modes of individual liability) the analyses did not
reveal any statistically significant patterns in the ICTY and ICTR sentencing practice. Therefore, it is not possible to make any conclusions regarding the relationship of individual
modes of liability and sentence severity.7 Finally, one of the main criticisms that could be
levied against the presented models concerns the fact that the tested variables represent only
very general aggregate categories and ignore further more detailed distinctions relevant in
sentence determinations such as differences among underlying offences of conviction or
individual mitigating and aggravating factors. As demonstrated in Chapter I, in particular with
respect to these detailed distinctions, the ICTY and ICTR sentencing case law is sometimes
confusing and disparate. Empirically, given the statistically speaking very limited sample size
(i.e. number of cases decided by the Tribunals so far), it was necessary to limit the possible
predictors included in the models substantially. In future, with more cases decided by the
Tribunals it would be possible to investigate further possible relevant aspects. By including only
the aggregate variables in the models these other potentially relevant mechanisms of sentence
determination remain hidden. Normatively speaking, the fact that mitigating factors indeed
mitigate a sentence and aggravating factors account for the sentence extension, as demonstrated
in the presented analyses, could be perceived as a basic necessary pre-requisite of any fair

---

6 However, in this case it is also possible, as discussed above, that differences in sentence severity among individual
categories of international crimes are connected to the assessment of the gravity in concreto, not to their legal
classification. The ICTY and ICTR judges always emphasize the importance of the gravity in concreto evaluation
in sentence determination. The presented analyses did not analyze the relationship between gravity in concreto,
legal classification of an act and sentence severity and this is one of the matter that could possibly be examined by a
future research on international sentencing.

7 As discussed above, this could have been caused by the methodological limitations of the study or by the lack of
empirical patterns in the ICTY and ICTR sentencing practice stemming from the lack of a principled approach of the
ICTY and ICTR judges to the issue of a relationship between modes of individual liability and sentence
severity.
sentencing system\textsuperscript{8} and to conclude on this basis that international sentencing is consistent might seem to be an overstatement. However, the fact remains that even on the basis of these general categories, the models accounted for a substantial amount of sentencing variations. Adding variables (that would capture more detailed distinctions among cases) would not reduce the amount of sentencing variations explained by the model.\textsuperscript{9} Consequently this thesis has demonstrated that the ICTY and ICTR sentencing practice seems to be generally consistent and predictable.

3. \textbf{METHODOLOGICAL LIMITATIONS}

As noted in the introduction to this thesis, the regression analysis and similar multivariate techniques are the required and optimal statistical tools to investigate the empirical reality of sentence determination. Since multiplicity of various factors arguably influences sentencing decisions, the multiple regression analysis with its ability to tease out the contribution of individual factors (each time given the other factors in a model) and their combination to sentence length is the most appropriate method. Given its model, this technique is able to answer the question: “What is, ceteris paribus, the contribution of a factor x to sentence length?” However, there are issues relating to the design of the empirical parts of this thesis that require discussion. One of the strong points of the presented analysis is the fact that the modelling was performed on the whole population, in a statistical sense, of cases decided by the Tribunals at the time of writing. Therefore, the results are actually based on the analysis of the entire population not a sample. However, the entire population of the ICTY and ICTR cases is not very numerous and therefore, the author was limited in her choices of modelled variables.

There is a vast array of legally relevant factors\textsuperscript{10} that arguably influence sentencing decisions in individual cases. Theoretically, also extralegal factors such as judges’ background or various characteristics of defendants could influence decisions on sentence severity in individual cases.


\textsuperscript{9} If variables measuring more detailed distinctions are added to the model and turn out to be statistically significant and improve the model, only coefficients of already included variables will change, the overall amount of correctly predicted sentences will not be reduced. For further discussion see the section on Methodological limitations below.

\textsuperscript{10} The legal factors are those considerations that should play a decisive role in the sentence determination. They are derived from the international law, the ICTY/ICTR Statute and the Rules and include factors such as nature of the crime, degree of responsibility or aggravating and mitigating circumstances. In contrast, the extralegal factors are factors not regulated or permitted by law. They should not influence the sentence decision-making in any way. Examples of these legally irrelevant considerations are e.g. ethnicity of the offender, professional background or personal characteristics of judges or political factors.
and in order to rule out these hypotheses, their influence should be tested as well. However, it was practically impossible to examine them all simultaneously in the presented analyses. With respect to extralegal variables, none of the empirical studies conducted so far has demonstrated any influence of such extralegal factors in sentence determinations at the ICTY and ICTR. The author, therefore, decided not to include them in the models. The limited number of legally relevant factors included in the models is primarily determined by the statistically small size of the population of the ICTY and ICTR cases. Therefore, the presented models tested only a limited number of variables based on general categories (such as category of crime or number of mitigating factors) rather than more particular but also relevant distinctions (such as underlying offences or specific aggravating or mitigating factors). Due to the limited ‘sample’ (in fact population) size, it was necessary to focus on this sort of ‘aggregate measures’. Therefore, the estimations could not reflect other possibly relevant dimensions of sentence determination such as influence of particular aggravating and mitigating factors on sentence length. It might be the case that in reality a single or several mitigating factors such as for example substantial cooperation with the Prosecution, remorse or family circumstances, to name few examples, has a stronger influence on sentence length while other mitigating factors influence the sentence less strongly. By including an aggregated variable reflecting only the number of mitigating factors in the models, these (possibly also) relevant differences were not examined.

Next, the selection of the number of independent variables to be modelled was also substantially limited by the small sample/population size. Therefore, the crime gravity, for example, was modelled by variables expressing category of convicted crime, number of counts, rank, the number of aggravating factors and modes of individual liability. Arguably, however, the crime seriousness could in reality be captured by a number of other variables such as underlying offences, a quantification of harm caused by a defendant such as number of victims or particular aggravating factors such as cruelty or premeditation. By omitting these dimensions of crime seriousness any effect does not emerge. As already discussed above, for example, the conviction for genocide popped out in the final model as the strongest predictor of sentence length. However, it might also be the case that in practice genocide counts are those with the greatest amount of harm and therefore entail the severest sentences. Since the model did not

---

12 As already indicated above, however, the judgments often do not include any indications on number of victims affected by a defendant’s crimes.
incorporate variables measuring these other dimensions of crime seriousness, the genocide variable might have picked up some of the effect of these omitted variables. As discussed further below, however, it is extremely difficult to quantify harm resulting from international crimes. For example, number of victims – the first obvious choice for such quantification – is sometimes not even indicated in the judgments or irrelevant for certain types of international crimes. How should harm be quantified in such cases? Arguably, more data and more research are needed to obtain improved measures of crime seriousness in order to fully and adequately reflect the richness of factors taken into consideration in sentencing decisions.\textsuperscript{13}

While this manuscript is one of the first studies examining jointly the ICTY and ICTR sentencing practice, the dataset is by necessity unbalanced. At the time of the presented empirical analyses, more cases have been decided by the ICTY (N=71) compared to the ICTR (N=40). Therefore, it was not possible to statistically test consistency in outcome at the ICTR separately. Only a descriptive overview of the ICTR sentencing practice was therefore offered. In a combined dataset, ICTY patterns might have dominated the results. Furthermore, the models presented in Chapter II and Chapter V slightly differ (the model presented in Chapter II included one extra variable on guilty plea and three separate variables on ordering, instigating and planning; while the model in Chapter V did not include a separate guilty plea variable and the three modes of liability were combined into one variable of Planning_Instigating_Ordering). In the process of learning and delving into the topic the author decided to adjust the original model in this way. The guilty plea did not seem to have an additional effect on sentence determination next to and above the other tested variables. Instigating, ordering and planning were combined into one joint variable for the following reasons: (i) theoretically, these three modes of liability are very similar in nature. The liability in all cases is derived from liability of a primary perpetrator and an order-giver, instigator and planner must act with the intent that the planned, instigated or ordered crime will be committed; (ii) empirically, it is often the case in the Tribunals case law that a defendant is convicted jointly as an order-giver/instigator/planner.\textsuperscript{14} In order to correct for these differences in the models, the model of

\textsuperscript{13} For a discussion of bias resulting from omitted or misspecified variables see A. Blumstein, J. Cohen, S.E. Martin and M.H. Tonry: Research on sentencing: The Search for Reform, Volume II, National Academy Press, Washington, 1983, at 141 et f..

\textsuperscript{14} All the convicted planners have also been convicted as order-givers/instigators. Out of 31 cases of convicted order-givers in 14 cases a defendant was convicted on the basis of combination of ordering and instigation and/or
Chapter V was tested on the ICTY data and vice versa. In these additional analyses it turned out that the explanatory power of these additional models remain the same (Adj. $R^2=0.593$ for ICTY data; Adj. $R^2=0.702$ for the combined ICTY and ICTR data) and the sets of statistically significant predictors was also almost identical. With respect to the ICTY data, the rank, conviction for crimes against humanity and number of counts were the strongest predictors of sentence length. The effect of conviction as planner/instigator/order-giver and number of mitigating factors had a borderline significance. Number of aggravating factors just missed statistical significance.\textsuperscript{15} With respect to the combined ICTY and ICTR data, the conviction for genocide, rank, number of counts, number of aggravating factors and number of mitigating factors were the strongest predictors. The conviction for crimes against humanity and conviction as order-giver just missed statistical significance. Consequently, the core set of significant predictors and explanatory power of the models remain the same. There are minor differences across the models in the effect of predictors and their statistical significance. Despite the limited number of cases in the sample compared to the number of predictors, the models in fact turned out to be remarkably stable.

Finally, it should also be noted that characteristics of defendants included in the study do not fully represent the whole population of individuals to be (potentially) convicted by the ICTY and ICTR. The cases still pending before the Tribunal’s (Pre)-Trial Chambers represent more complicated and arguably also more serious cases, in the sense of seniority of the accused and relating crime base/crime gravity, compared to the already finished cases. Once these cases are decided and included in the sample, the effect of these cases might also influence the results. However, given the fact that these cases are not so numerous (at the ICTY 24 cases and at the ICTR 20 cases not included in this study are pending) and the fact that arguably the sentencing practice at the Tribunals is rather stable now (judges refer to previously decided cases and compare sentencing outcomes) it is not likely that their inclusion in the data would substantially alter the results.

\textsuperscript{15} Planning. There have been 17 cases of instigators, only in 5 of them instigation alone formed the basis for a conviction.

\textsuperscript{15} In this model, however, the residuals do not seem to be normally distributed which might have affected the accuracy of the model.
4. IMPLICATIONS – INTERNATIONAL SENTENCING GUIDELINES?

The discussion of consistency and disparities in any sentencing regime is closely connected to a discussion of sentencing guidelines. In domestic jurisdictions many initiatives have been proposed and implemented to structure sentencing decision-making and limit discretionary powers of judges in order to make sentencing fairer and more transparent. Also at the international level, several commentators called for international sentencing guidelines. Sloane argued that ‘the adoption of genuine sentencing guidelines by international tribunals would be a significant step toward rationalizing the sentencing process’ Beresford noted that ‘[u]nstructured discretion allows in the personal preferences of a judge and may permit discrimination, individual idiosyncrasy and other irrelevant influences to encroach upon the discretionary framework provided by the Statute and the Rules. The establishment of sentencing guidelines would avoid such occurrences...’ Several questions could be asked in this respect. Is it indeed necessary to develop international sentencing guidelines and if so, is it even possible given the complex nature of international crimes?

4.1. Are international sentencing guidelines necessary?

Empirically, as demonstrated by this manuscript, it seems that international criminal judges have developed a rather consistent sentencing practice whereby a substantial amount of sentence variation could be predicted on the basis of several legally relevant factors. As already discussed

---

above, in the presented sentencing models very general variables measuring broad legal categories were included because the limited sample size did not allow conducting a more detailed examination. On the basis of these very general factors already around 70% of sentence length variation is accounted for. Therefore, international sentencing as practiced at the ICTY and ICTR can to a large extent be predicted by legally relevant factors and seem to be as predictable as domestic sentencing practices with much more stringent sentencing guidance. It is a different matter that many questions relating to the role of individual legal factors in sentence determination remain unresolved and should be investigated by a further empirical research of the international sentencing practice.

Normatively, given the limited guidance in the positive law, it could be argued that some sort of sentencing guidance would make international sentencing more transparent, rational and easier to review. Indeed as demonstrated in Chapter I, the sentencing determinations at the ICTY and ICTR are not examples of a crystal clear and at face value understandable sentencing system. There is a lack of transparency in the sentencing argumentations of the ICTY and ICTR judges and it is not clear what punishment is ascribed to what conduct. In this respect, it would be desirable to develop more structured and transparent sentencing. At a domestic level, several tools have been implemented to develop consistent sentencing standards and to address alleged disparities in sentencing such as mandatory minimum sentences, determinate sentencing law, guideline judgments or numerical sentencing guidelines. For example, numerical sentencing guidelines as implemented in many jurisdictions in the United States, seek to channel judicial discretion in individual cases by providing a recommended disposition for each offence (e.g. probation, fine, prison) and a proposed sentence length or range when a prison term is recommended. Broadly speaking, numerical sentencing guidelines typically consider a particular level of crime severity and the criminal record of the offender as the basis of a computation of a recommended sentence allegedly promoting uniformity and fairness. Consequently, under the numerical guidelines consistency is achieved by requiring judges to conform to narrow sentence ranges and to justify any departures from these ranges by finding ‘substantial and compelling reasons’. Here consistency is defined in terms of consistency in outcome. ‘Thus, courts are directed to impose a sentence within a narrow range of sentence length and no methodology is provided by which to arrive at the final sentence. Consistency is achieved by ensuring a

sufficiently high number of sentences fall within a narrow presumptive range.\textsuperscript{20} This sort of numerical guidelines has however been subjected to fierce criticism on many fundamental grounds such as (i) policy-related (sentencing guidelines unduly narrowed judicial discretion and shifted the discretion to prosecutors); (ii) process-related (in practice, they are being circumvented by prosecutors and judges); (iii) ethics (sentencing guidelines forced key decisions behind closed doors and fostered hypocrisy in a system); (iv) technocratic grounds (the guidelines are usually too complex and hard to apply accurately); (v) fairness (sentencing guidelines actually lead to unfair sentences because only the offence or offence behaviour and criminal record is taken into account not other circumstances pertinent to a sentenced individual); or (vi) outcome and normative grounds (sentencing guidelines have not in fact reduced sentencing disparity and sentences based on these guidelines are too harsh).\textsuperscript{21} An alternative to this rather rigid approach has been favoured in parts of Europe, such as Sweden, Denmark or the United Kingdom. In these countries, sentencing is regulated by more flexible rules that basically provide detailed and structured methodology for courts to follow. In these systems consistency in approach is promoted and achieved ‘by structuring the decision making of a court at sentencing, effectively by creating an algorithm to be followed. The underlying logic is that, if two courts sentencing two different cases of, say robbery impose different sentences, having followed the same sequence of steps to determine sentence, the difference between the dispositions is likely to reflect legally relevant factors.’\textsuperscript{22} This less restrictive approach to sentencing regulation has not led to so much criticism compared to the strict numerical sentencing guidelines.\textsuperscript{23} Theoretically, therefore, there are many options how to channel sentencing discretion of international judges. Some suggestions with respect to ‘a narrative guidance’ and a development of a more transparent and consistent approach to sentencing at the ICTY and ICTR were discussed in Chapter I. The question remains whether initiatives leading to more rigid sentencing structures such as numerical guidelines or sentencing


\textsuperscript{22} J.V. Roberts, supra note 20, at 15.

\textsuperscript{23} G. Mackenzie, supra note 21, at 81.
grids are even possible at the international level especially taking into account the complex character of international crimes.

4.2. Are international sentencing guidelines possible?

Setting aside practical considerations as to how such measures would be drafted and enforced at the international level, sentencing guidelines ‘stricto sensu’ in the sense of numerical grids applied in many US jurisdictions or mandatory sentencing minima and maxima (hereinafter ‘sentencing guidelines’) are arguably not feasible and/or optimal way for structuring international sentencing. In practice, sentencing guidelines have either been descriptive (i.e. based on statistical characterization of previous sentencing practices) or prescriptive (i.e. results of policy decisions about appropriate punishment not connected to previous sentencing practice).

With respect to the former, it would be difficult to come up with the descriptive sentencing guidelines based on the international sentencing practice since this is still very limited and fragmentary. Despite many developments in the field of international criminal justice in the last decade, the data on sentences at the international level are still scarce. Another problematic issue of such descriptive sentencing guidelines relates to the fact that even if there were enough data, the sentencing guidelines developed in this way would actually freeze a further development of the international sentencing practice since the future sentencing would be bound to the early practices of international tribunals.

These considerations lead us to the question of prescriptive sentencing guidelines that should reflect policy choices regarding the appropriate punishment for particular offences. Assuming that individual responsibility and individual sentencing (based on the domestic concepts) would be retained at the international level, it is still extremely complicated to come up with pre-determined sentencing standards, such as recommended sentences or sentencing ranges. Arguably, theoretically developed international sentencing guidelines should reflect the peculiarities of international crimes. It is however extremely difficult to ‘squeeze in’ and in advance quantify these peculiarities of international crimes and come up with pre-determined

---


sentencing ranges. On the one hand, the sentencing ranges for international crimes should be broad enough to accurately reflect diversity in seriousness of individual international crimes and all the complexities relating to culpability of their perpetrators. On the other hand, since the main aim of these initiatives is to structure the sentencing discretion and enhance consistency of sentencing, the sentencing ranges should not be overly broad in order to have any practical significance and limit judges’ discretion.

In order to come up with ‘international sentencing guidelines’ it would be necessary to establish a clear system of sentencing objectives. Adopting a particular objective has important implications for the development of sentencing guidelines – it determines what are the focal points and other relevant factors in sentence determination. This question in itself is extremely complicated as demonstrated by the scholarly discussion of this topic touched upon in Chapter I. The English philosopher H.L.A. Hart provided a useful framework for consideration of the normative goals of punishment. He noted that debates about the philosophy of punishment at a domestic level often tend to be unnecessarily complicated and proposed that we should devote separate attention to the following three distinct questions: (i) The general justifying aim – What is the general justification of the social institution of punishment?; (ii) The question of liability – Who is to be punished?; and (iii) The question of amount- How much?.

In this respect, several punishment theories could be simultaneously employed each at a separate level of analysis to serve as normative goals of international sentencing. Focusing on the amount of punishment, let us assume that we would, as the ICTY and ICTR at least in a rhetorical sense did, employ retribution as the underlying theory to determine how much punishment would be appropriate for individual offences. Under this normative goal, the punishment should fit the crime and should be proportionate to the seriousness of the committed crime – a defendant should receive his ‘just deserts’. Retributivists’ main concern is the coherence of this offence severity scale – the punishment should be proportional to the seriousness of the crime (‘cardinal proportionality’) and crimes and their punishments should be rank ordered relative to each other in terms of their seriousness and severity (‘ordinal proportionality’).

26 A. Blumstein et al., supra note 24, at 48.
domestic level individual offences are rank ordered in terms of their relative gravity and sentencing guidelines provide multiple ranges of sentence length corresponding to different levels of crime seriousness.

To come up with ‘a crime gravity=sentence severity equation’ was an extremely difficult task at the domestic level – to assess harm of a given category of crime and relate it to the injuriousness of all the other crimes is very complicated. Andrew von Hirsh has developed a theory of proportionality applicable to domestic offences focusing on ‘the harmfulness’ of criminal conduct. In order to establish some parameters of ordinal proportionality he focused on the question of what generic interests are violated or threatened by a standard case of an offence: physical integrity, material support, freedom from humiliation or privacy and autonomy. To assess the crime seriousness, however, other issues must also be addressed. He proposed a four-dimensional assessment of offence seriousness: the first dimension determines the interests violated. In the second dimension there is a preliminary quantification of the effect of a typical case on a victim’s living standards. In the third dimension account is taken of the culpability of the offender. And finally, there may be a reduction in the level of seriousness to reflect the remoteness of the actual harm. This scheme, however, is difficult to transfer to international crimes since (i) arguably a majority of international crimes would be ranked on the highest level of this seriousness scale since they usually target a victim’s life and/or physical integrity while profoundly affecting victim’s living standard; (ii) these crimes violate interests not only of individual victims but target entire communities or groups of individuals and (iii) arguably the most culpable perpetrators are often far remote from ‘the actual harm-causing behavior’.

International crimes are usually offences entailing a large number of victims, committed as part of a large-scale campaign and implicating many individuals with very different tasks in the whole ‘criminal enterprise’. They are committed through acts that resemble crimes under domestic criminal law. However, what distinguishes murder or torture as an international crime from ordinary murder or torture is the context in which they are committed – they must be

---

29. A. von Hirsch, N. Jareborg: Gauging Criminal Harm: A Living Standard Analysis, 11 Oxford Journal of Legal Studies 1 (1991) 1-38, at 19; Hirsh explains that the assessment of ‘standard harm’ is not universally applicable but culturally relative and it will differ in different cultures with different underlying values.

30. A. Ashworth, supra note 28, at 109.

31. In the Hirsch-Jareborg scheme, most international crimes would be ranked as ‘gravely harmful’ arguably intruding into interest required for subsistence of victims (Level 1 in their four-level scheme of seriousness). See Hirsch, Jareborg, supra note 29, at 17.
committed within an attack targeting broader groups of people, not only a single individual. Therefore, there is not only ‘harm’ or ‘living standard’ of individual victims to be taken into account but also ‘the aggregate harm’ of targeting any civilian population or destroying ethnic or religious group. At the international level, therefore, even more considerations enter the equation of gauging crime seriousness relative to other crimes and the appropriate level of punishment. Consider a crime of homicide/murder as one of the most serious crimes entailing the most serious penalties such as life imprisonment at a domestic level.\(^{32}\) If a single murder already entails the severest penalty, what about murdering tens or hundreds of victims? At the international level it could be that a single murder forms the basis of conviction but it could also be that an individual is found responsible for taking lives of hundreds or even thousands of victims. What is the proportionate punishment in such case? How much more time in prison should cost the lives of hundred victims more? And even if an individual is punished for a single murder as an international crime, this crime must be part of a much larger campaign and, as noted above, the contextual elements of the international crime of murder/killing should therefore be reflected in sentence ranges as well. How, however, would we in terms of sentence length express his/her genocidal intent to destroy a whole ethnic or religious group? Is this intent worth more punishment than knowledge that his/her crimes were committed as part of a large scale attack against any civilian population?\(^{33}\) And if so, how much more?

Next to the assessment of ‘objective harm’ caused by individual offences, it is also necessary to foresee and quantify different degrees of culpability of perpetrators of international crimes. Many perpetrators with varying degrees of culpability get involved in international crimes and rigid sentencing guidelines can hardly reflect such a complexity. At the domestic level the assessment of culpability is often linked to the degree of intent and a lack of ‘sentencing excuse’ such as partial duress, provocation or partial mental disability.\(^{34}\) At the international level, things get again more complicated. Since international crimes are distinguished by their collective and systematic character, it is arguably also necessary to assess a defendant’s role in the overall system to evaluate his culpability. These crimes are often state sanctioned and characterized by mass involvement of both military functionaries as well as civilians. Many otherwise law-abiding citizens get involved in a period of collective violence and commit their crimes on direct

---


\(^{33}\) These ‘contextual’ concepts are particular for international crimes and the ICTY and ICTR judges have not tried to address their implications for international sentencing.

\(^{34}\) Hirsh, Jareborg, supra note 29, at 3.
orders, instigation or with acquiescence of political or military authorities.\(^{35}\) Since international crimes are often committed by masses in an environment where arguably criminal conduct becomes ‘normalized’, criminologists studying international crimes labelled international crimes as ‘crimes of obedience’\(^{36}\) and argued that international crimes are a different type of criminality than ordinary crimes. They are crimes of conformism rather than crimes of deviance. In times of collective violence many otherwise law-abiding citizens get involved and commit international crimes simply for the fact that ‘everybody else was doing it too’. Perpetrators of international crimes violate international rules but often act in conformity with the social norms and values which are prevalent in that particular state at that particular time.\(^{37}\) Therefore, it has been argued that sentences for international crimes should reflect both the environmental forces as well as the autonomy and individual agency of the people involved. Many perpetrators operate in a social environment in which international crimes are considered necessary and legitimate.\(^{38}\) According to the criminologists perpetrators of international crimes can be divided into three broad categories: conflict entrepreneurs who are instrumental for the development of system criminality, leaders who transfer their illegal orders and actual killers who act following such orders or out of conformism - often the most numerous category. As noted by Drumbl ‘these groups represent descending levels of moral blameworthiness…[C]onflict entrepreneurs are the most culpable according to the standards adopted by traditional criminal law, namely intentionality of action. They are followed by other leaders and killers…’\(^{39}\) However, many of these theoretically less culpable ‘conforming criminals’ commit extremely serious offences. On the other hand, the most culpable perpetrators – conflict entrepreneurs- are usually far away from the actual crime scene and it is very difficult to prove their criminal involvement. If convicted, it is sometimes (due to the lack of evidence) as ‘mere’ accomplices on the basis of ancillary modes of liability such as aiding and abetting that are generally considered less serious


than the actual hands-on perpetration.\textsuperscript{40} Therefore, there is a paradox of the most culpable perpetrators far removed from the actual harm-causing behaviour. Furthermore, theoretically a conflict entrepreneur and an ordinary soldier can both be convicted of committing a genocidal murder, however, given the systematic character of international crimes, their culpability level fundamentally differs. Consequently, sentencing ranges for international crimes would have to be extremely broad to capture all these variations and complexity of international crimes.

Therefore, there is a multiplicity of elements and unanswered questions with respect to ‘the crime gravity - sentence severity equation’ for international crimes. These are all extremely challenging issues and much more theoretical and empirical research is needed to address them. Many authors have theoretically argued that there should be a hierarchy of international crimes with respect to sentencing and that general ranges of penalties for underlying offences of international crimes should be established.\textsuperscript{41} However, how these should be translated to practice has never been discussed. How to establish a sentencing range for genocidal killing when in practice this offence can be committed by killing one person but also thousands? How then would a sentence for the crime of extermination as a crime against humanity entailing killing on a large scale relate to the sentence for genocidal killing? And many more similar questions can be identified with respect to developing pre-determined sentencing ranges for international crimes. Therefore, arguably it is extremely difficult to come up with sentencing ranges unless they are sufficiently broad to reflect all possible variations and considerations illustrated above. Then, however, the practical utility of such a broad sentencing ranges could be questioned.

The international criminal justice system faces many paradoxes with regard to sentencing international crimes and only few of them were briefly discussed above. It is not the aim of this thesis to develop a penal regime and philosophy of penalty for international crimes. The ICTY


and ICTR case-law, save for few very limited exceptions\textsuperscript{42}, did not elaborate on the specific nature of international crimes and its relevance for sentencing. As noted above, judges transferred domestic concepts and principles of sentencing law into their own narrative without any further ado. Therefore, sentences for international crimes are actually very comparable to, and sometimes even more lenient than, domestic sentences for ordinary crimes.\textsuperscript{43} The paradox then arises that a perpetrator of a multiple murder and harsh mistreatment of many victims is actually punished less or comparably to an individual convicted of a single murder at the domestic level. How could this paradox be explained and is this indeed what we, the international community, want is the question that remains to be answered. At the domestic level sentencing guidelines were developed to deal with relatively simple cases of single instances of crime such as murder or robbery. Even for these ‘single-incidents’ crimes it was extremely challenging to develop clear and easy to follow guidelines that are accepted and conformed to by judges.\textsuperscript{44} As briefly addressed above, the international crimes are usually extremely complicated, bulk crimes consisting of multiple offences and multiple perpetrators take part therein. It is very difficult to imagine that ‘simple enough’ sentencing guidelines that would be followed by judges would be able to capture all the complexity of international crimes.

5. \textbf{Further Research}

This thesis is one of the few studies investigating the empirical reality of international sentencing. It focused on consistency of international sentencing as practiced by the ICTY and ICTR and tried to identify any consistent patterns in their sentencing jurisprudence and practice. Since sentencing of international crimes still remains a rather neglected phenomenon in scholarly work, there are still plenty of unexplored issues to be addressed. Many of these issues have been already touched upon in the above text. This section highlights some possibilities for further empirical and theoretical research in this field.

Follow-up empirical research should focus on more detailed modelling of international sentencing and come up with more elaborate measures of crime gravity and individual circumstances of defendants. With more available data such modelling will bring further

\textsuperscript{42}A. Smeulers, B. Hola, supra note 2, at 188-189.
insights into mechanisms underlying the sentencing of international crimes. Furthermore, with more sentences handed out by the international tribunals it would also be desirable to test possible effects on sentence length of other legal (but also extra-legal) variables to rule out and/or confirm their possible influence on sentence determination. The examples thereof constitute proceedings-related factors (such as quality and amount of evidence, the fact that it is a joint trial of several defendants or of a single defendant, guilty plea\textsuperscript{45} or time of guilty plea) or defendant-relative variables (such as age, ethnicity, gender\textsuperscript{46}) or decision-maker-relative variables (such as judges background, legal system they come from or gender).\textsuperscript{47} In this respect, also qualitative research methods such as observations or interviews with international criminal justice participants could bring further insights into the dynamics of judges sentencing decision making. Since quantitative methods cannot capture many nuances of the decision making process such as social interactions or subjective attitudes of individual decision-makers and their influence on decision making, the qualitative research techniques are particularly useful tools in this respect. Interviews with international judges can bring further insights into the concepts such as crime gravity or culpability of perpetrators and might also help in creating statistical models that more accurately reflect the process of sentence determination. Further empirical research should also investigate the practice of early releases at the ICTY and ICTR and how this affects the ‘effective’ sentence actually served by international criminals. The ICTY and ICTR Statutes provide for the possibility of early release (according to the law of a state where a sentence is served) after a decision of the Tribunal’s President. There has been no empirical enquiry describing the actual execution of international sentences and research in this area would bring many useful insights into the ‘post-conviction’ stage of the international penal practice.\textsuperscript{48} In a broader perspective, in order to assess conclusively the unity or fragmentation of the international sentencing system, sentencing of other international and internationalized tribunals should be investigated in more depth. What are the relationships and influences, in

\textsuperscript{45} The guilty plea as a variable was included in the ICTY model and did not end in the final model as a significant predictor of a sentence length. When the model with a guilty plea was tested on the combined ICTY and ICTR data, its coefficient also was not statistically significant. One of the reasons for this phenomenon might be that the effect of guilty plea was subsumed in the overall effect of the variable measuring number of mitigating factors. Therefore, in the future studies when it will be possible to test effects of individual mitigating factors on sentence length, it might be that the separate effect of guilty plea on sentence length will be found.

\textsuperscript{46} However, so far there have been only two women convicted and sentenced by the ICTY and ICTR.

\textsuperscript{47} Some of these extralegal variables were already tested in some other studies and as noted above, never turned out to statistically predict sentence length.

\textsuperscript{48} For overview of problems connected to the execution of sanctions at the international level see J.C. Nemitz: Execution of Sanctions Imposed by Supranational criminal Tribunals in R. Haveman, O. Olusanya (Eds): Sentencing and Sanctioning in Supranational Criminal Law, Intersentia, Antwerpen, 2006, 125-144.
terms of sentencing for international crimes, among the international criminal courts and tribunals that have emerged in the past decade? In view of the principle of complementarity, one of the governing principles at the International Criminal Court, influences between domestic courts/practices dealing with international crimes and their international counterparts also need to be explored.

Theoretical inquiries should address such concepts as the principle of proportionality in international sentencing and how it differs from proportionality as applied in domestic jurisdictions. Some scholars already touched upon the subject.49 D’Ascoli, for example, argued that the principle of proportionality in international sentencing should not be focused on harm caused by an offence but on culpability of a perpetrator. The proportionality equation would thus be ‘culpability of a defendant = severity of sentence’. The adjusted formula, according to D’Ascoli, should be that the imposed penalty should match primarily the degree to which the offender is culpable for the crime and his personal circumstances.50 To this it can be added that the context in which international crimes are committed and the particular role of a defendant within this context should then play a primary role in the assessment of his/her culpability. Taking into account criminological insights, briefly discussed above, in order to reflect the systematic character of international crimes the assessment of culpability of a perpetrator should take into account the overall circumstances and socio-political context in which international crimes are committed. In order to reflect properly the role of various perpetrators of international crimes it is necessary to distinguish between those who (i) exert influence within a society/criminal organization and are instrumental for the reversal of societal and moral norms of a given society creating a social reality in which international crimes become acceptable; and (ii) those who are influenced to participate in international crimes by a reversed social order and conformism/obedience.51 These considerations are related to much broader problems addressing the questions of linking international sentencing and prosecution to the aetiology of international crimes. The current practice of the international tribunals has been criticized as not sufficiently

50 D’Ascoli, supra note 41, at 293.
51 A. Smeulers, B. Hola, supra note 2, at 203; For an opposing view see J.D. Ohlin, supra note 48, where the author argues that from the normative perspective the proportionality based on offences committed should be prioritized against taking into account a position of a particular defendant in a conflict and ordinal proportionality among different crimes of different defendants.
addressing the peculiarities of international crimes and simply transposing concepts from domestic prosecution and sentencing. More research is needed in this respect to further promote the understanding of international crimes as collective criminality and its peculiarities with respect to punishment and sentencing individual perpetrators.

6. **Concluding Remarks**

The international sentencing practice of the ICTY and ICTR has been largely criticized in academic discourse. Various scholars have labeled the ICTY and ICTR sentencing as irrational, akin to a lottery system or inconsistent. This thesis has demonstrated that this heavy criticism is overstated. Consistent empirical patterns have emerged in the Tribunals’ sentencing case law and practice and the ICTY and ICTR sentences are to a rather large extent predictable. The presented empirical analysis has proved that a set of legally relevant factors derived from the general sentencing principles, which are recurrently emphasized by the ICTY and ICTR judges, has a consistent and predictable effect on sentence length.

Therefore, it seems that the main problem of the ICTY and ICTR sentencing may not so much be a lack of consistency but a lack of transparency and clarity of sentence determinations. ICTY and ICTR judges do not properly explain how sentences in individual cases were determined: it is not clear from the judges’ reasoning how they came up with a final sentence and why the final sentence is appropriate in a given case. Judges very rarely indicate what weight is ascribed to individual sentencing factors and how do they sum up to form the final sentence. Given the very rudimentary positive law framework, it is, however, extremely important to develop (for the sake of defendants but also other international criminal courts and tribunals) clear sentencing case law. As pioneers of the re-established international criminal justice system after the Cold War, the ICTY and ICTR are setting a lot of important precedents and sentencing jurisprudence is by necessity one of them. Since the sentencing argumentation at the IMTs in Nuremberg and Tokyo was almost non-existing, even more attention should be dedicated to sentencing by the Tribunals because their case law will serve as a reference for other international criminal courts and tribunals\(^5\), including the permanent International Criminal Court. International judges should try more to present comprehensive and transparent sentence narrative so that not only

---

\(^5\) See for example Art 19 of the Statute of the Special Court of SierraLeone (SCSL) that dictates the SCSL judges to have recourse to the practice of the ICTR. Similarly, Art 24 of the Statute of the Special Tribunal for Lebanon refers judges to have recourse to international practice regarding prison sentences.
defendants and the Prosecution in individual cases but also the general public are aware of why and how the amount of punishment in a particular case is as it is. The consistent empirical patterns discovered in this thesis are in some cases not obvious from the judges’ narrative and this is one of the main flaws of the ICTY and ICTR sentencing practice as it now stands. If the ICTY and ICTR judges were clearer in their sentencing argumentation, the majority of the heavy criticism might have been prevented and the sentencing practice of the ICTY and ICTR would have probably gained more legitimacy and support. Regarding the actual consistency of international sentencing the findings of this thesis seem to be encouraging. Despite all the criticism leveled against the Tribunals’ sentencing practice, this thesis demonstrated that empirically it is possible to identify legally relevant and consistent patterns in the sentencing narrative and outcomes of the ICTY and ICTR. International sentencing as practiced by the ICTY and ICTR does not seem to be a random system with no order whatsoever or ‘a game of a Russian roulette’ but on a general level a consistent and predictable practice.