Chapter 4

GUIDELINE AND SAMPLE SENTENCES COMPARED

Introduction

How do sentences given by the American public compare to those prescribed in the sentencing guidelines? Answering this question in some detail is the major goal of this chapter.

There are several good reasons for making these comparisons. First, the guidelines were written in response to a legislative mandate requiring the Commission to be guided in some significant measure by the relative seriousness of criminal offenses.[1] Serious crimes were to be given harsher sentences and crimes of lesser seriousness were supposed to be treated more leniently. The several social science studies of public beliefs about the seriousness of crimes reviewed in Chapter 1 have shown that the public has well structured views about relative seriousness: there is a fair amount of consensus on the seriousness ordering of crimes. How well was that consensus transformed into respondents' recommended sentences?

Second, there is the issue of public confidence in the fairness and equity of the criminal justice system. Although there is no strong reason why guideline sentences should be patterned precisely after public preferences, it would be a matter of concern were the two to be found far apart. A criminal justice system risks losing its legitimacy if it provides sentences that are much harsher or much more lenient than what the majority of its constituents believe to be just.

For these reasons we can expect some degree of correspondence. However, there are also good reasons to expect differences between guideline sentences and those desired by the public. Such differences are bound to occur because sentencing also reflects other goals besides maximizing legitimacy. Among many other considerations, some involve possible deterrence effects of sentencing, fiscal implications of the costs of prosecution and imprisonment, and the ease or difficulty of running the courts.

Accordingly there are no clear expectations about how closely comparable guideline sentences and those given by the public should be. The degree of acceptable comparability may be largely a matter of judgment. In addition, there are also technical considerations. As shown in Chapter 3, Americans cannot be considered to be of one mind on sentencing for almost all of the crimes studied. There were some who were in favor of each of the alternatives in sentencing offered for almost every crime. The critical issue is how best to summarize what were central tendencies in sentencing, a difficult decision when average and median values for most crimes were far apart.

In addition, there is the issue of how to measure comparability. A stringent definition is one in which comparability is defined as identical sentences. A less demanding definition of comparability requires only close rank orderings of sentences for crimes, in which it is possible for the public views and guideline sentences to rank crimes the same way but differ systematically in the sentences imposed. Several approaches to measuring comparability will be considered, from among which the reader can choose.

---

[1]The authorizing legislation is contained in 28 § U.S.C. 991(b)(2).

Finally, comparability can be explored at different levels of aggregation. At the finest level, the sentences given by individual respondents to specific vignettes can be compared to the calculated guideline sentences for the same vignettes. This comparison assesses the extent to which individual respondents agree with the guidelines in giving sentences to specific vignettes. Aggregating respondent sentences by calculating mean or median sentences provides another way of making comparisons in which guideline sentences are contrasted with the central tendencies found in public opinion. Several levels of aggregation will be explored in this chapter.

Calculating Guideline Sentences

In essence, the guidelines consist of a set of rules for arriving at sentence ranges for persons convicted in the federal courts. The guidelines take into account both the crimes of which the defendant was convicted and the actual nature of the criminal conduct by assigning a "base offense level" (a number) that serves as a starting point in assessing the seriousness of an offense. This base offense level can be increased or decreased based on the circumstances of the case. The factors that modify the base offense level (such as use of a weapon, presence of more than minimal planning, or amount of loss) are enumerated in the guidelines.

A base offense level, modified by the specified circumstances and other general adjustments (such as role in the offense) forms one axis of a table used to determine sentence ranges. The other axis reflects the defendant's criminal history as expressed in one of six categories. The point at which the adjusted offense level and criminal history category intersect in the table determines a range of sentences for an offender. Ordinarily a federal judge must choose a sentence from within the guideline range unless the court can identify a relevant factor not covered in the guidelines; in this case, the judge may depart from the guideline range and must provide a reason for so doing.

The guidelines are carefully documented and permit a person with knowledge of the relevant statutes and the details of an offense to compute the guideline sentence range recommended. For example, by following the guideline structure carefully, one can calculate that a sentence range of 46 to 57 months would apply to a person with no previous criminal history who was convicted of extorting $19,000 from a victim by threatening to kill someone in the victim's family.

The fact that there are parallels between the vignettes and the guidelines is, of course, no accident but part of the vignette design strategy. Many of the dimensions built into the vignettes paralleled provisions in the guidelines.[2] For example, the guidelines indicate that sentences for drug trafficking offenses should vary in specific ways according to the drug in question, the amount of the drug involved, the role played by the defendant in the transaction, and the presence or use of weapons associated with the transaction. Accordingly, the vignette design for drug trafficking calls for vignettes to vary the type of illegal drug, the amount of drugs involved, the role played by the convicted person in the drug trafficking, and weapon use. Most important of all, the parallels make it possible to calculate a guideline sentence for each of the vignettes used in the study.

---

[2]Some of the dimensions used in the study are not recognized in the guidelines as grounds for sentencing decisions. For example, the vignettes vary the gender of the offender but the guidelines do not recognize gender as a factor in sentencing.

The translation of the crimes and dimensions used in this study into recommended sentences was done by staff members of the U.S. Sentencing Commission. Their familiarity with guidelines made the translation of the vignette dimensions into guideline sentences as faithful as possible. However, the translation was not without problems. First, the guideline definitions are more precise than many of the vignette descriptions. For example, the guidelines include an adjustment for "aggravated role" if a defendant's criminal behavior included organizing or managing criminal activity involving five or more participants. However, the corresponding vignette dimension did not specify the number of persons managed by the defendant; it was unclear whether this substantial adjustment should have been applied.

Second, some offense details included in the guidelines affect sentences by way of judicial discretion at the time of sentencing. For example, the guidelines permit the judge in conservation and wildlife cases to depart from the guideline sentence range when the quantity or seriousness of damage is "not adequately" measured by the guideline sentence. Consequently vignettes describing such crimes can be given guideline sentence ranges, but actual sentences imposed can often fall above those ranges.

As a consequence of the two ambiguity factors, the mapping of guideline sentences on vignettes was not always easy and in some cases impossible. It is important to understand that these translation difficulties applied mainly to dimensions that modified the two major features of the vignettes, the crimes committed and the offenders' prior records, the main determinants of guideline sentences. The crimes committed and the offenders' prior records together accounted for 92 percent of the variation in guideline sentences.[3] In short, for the most important features of each vignette, the assignment of guideline sentences was accomplished typically with little ambiguity. However, for some of the minor other dimensions incorporated into the vignettes, the translation of levels into sentence enhancements could not be accomplished with great confidence.

The guideline recommended sentences consist of a range of acceptable sentences. The midpoint of the guideline sentencing range for each crime description in a vignette was used to represent the recommended guideline sentence and attached to the crime as described in each vignette.[4] Accordingly, the extortion crime used as an example above was given a guideline sentence of 51.5 months or 4.2 years. Each of the vignettes studied was given a guideline recommended sentence that varied according to the crime described, the levels of crime dimensions used and the previous record of the convicted offender. For example, the calculated guideline sentences for vignettes involving extortion crimes ranged from .58 to 10.3 years, the differences among sentences reflecting varying amounts of money extorted and the previous record of the convicted offender.

For the crimes studied, there are no death sentences recommended in the guidelines. As discussed in Chapter 2, life sentences given by our respondents were translated as 70 years and death sentences as 100 years, translations which are used in most other chapters of this report. However, in this chapter we adopt a different convention. Because very long sentences can influence comparisons very heavily, in this

---

[3] This estimate was calculated by regressing the guideline sentences for each vignette on the crimes and prior records of each vignette, entered as dummy variables. The resulting R^2 was .92.

[4] Using the lower or upper boundaries of the guideline sentence ranges was also considered. Preliminary analyses indicated that using these alternatives did not affect results markedly. For example, using the lower boundary typically in regression typically affected the intercept values but not the regression coefficients.

chapter we have recoded respondent sentences in excess of that amount to 39.2 years.[5] Of course, it should be borne in mind that a reasonable argument can be made that any vignette to which a respondent gave a death sentence should really count as a disagreement with the guidelines.

For each of the vignettes used in the study, we calculated a guideline sentence taking into account the crime described in the vignette, the previous record of the convicted person and the features of the crime depicted in the dimensions shown using the scoring system implicit in the guidelines. It should be borne in mind that the calculated guideline sentences are based only on the information contained in the vignettes and that sentence enhancements were sometimes translated with some uncertainty.

Actual cases coming before the courts are much richer and more detailed. Guideline punishments for concrete cases would take into account features of cases which were not incorporated into the design of vignettes. Furthermore, the guidelines recommend a range of sentences of which we have taken the midpoint. Accordingly, the results shown in this chapter are based on a comparison of respondent sentences with guideline sentences as calculated with some uncertainty, believed to be minor in nature.

Analysis Strategy

The data set permits several kinds of comparisons between guideline sentences and those made by respondents. At the most disaggregated level, comparisons between guideline and respondent sentences can be made for each of the close to 70,000 vignettes. At the vignette level, the comparisons are of how close the 70,000 respondent sentences are to the guideline sentences for those vignettes. At the most aggregated level, we can compare how close respondent sentences come to guideline sentences for Crime Types consisting of broad classes of crimes, using the mean or median sentences given by respondents. There are 20 Crime Types used in the study. Comparisons on this level indicate the extent to which, say, the mean or median guideline sentence for larceny crimes compares to the mean or median sentence given to those crimes by respondents. A third level of intermediate aggregation centers around the 73 Crime Examples, each being a concrete instance of one of the Crime Types. A fourth approach is to consider each Crime Example as modified by the prior criminal record given to the offender. There are 175 combinations of unique examples and prior crime record which can be used. A fifth approach is at the level of individual respondents and calculates the extent to which individual respondents' orderings of the vignettes they judged correspond with the guidelines' orderings.

Because of the considerable amount of inter-respondent variability in sentencing (as shown in Chapter 3) less comparability can be expected at the vignette level, the greatest amount of comparability at the level of Crime Types and an intermediate degree of comparability at the Crime Example level. Using means or medians simply reduces inter-respondent variability.

Each mode of analysis is useful. For example, if it is assumed that a citizen's level of satisfaction with the federal criminal justice system is related to the differences between how he or she would sentence criminals compared to the guideline sentences, then the study of the distribution of satisfaction is well served by the comparisons at the level of individual vignettes. However, if we are concerned with whether

---

[5] This is the life expectancy of persons who are at the average age of felons at the point of conviction. This number is also used by Commission staff when a numerical equivalent is needed for life sentences.

the guideline sentences minimize differences with individual citizens in the aggregate then comparisons at more aggregated levels of Crime Types or Crime Examples may be more appropriate.

Comparisons at each level of aggregation will be presented in this Chapter. We start with comparisons at the level of individual vignettes and end with analyses at the level of Crime Examples.

Distributions of Guideline and Respondent Sentences Given to Vignettes

After removing the vignettes rated by outliers, as discussed in Chapter 3, there remain 68,712 vignettes for which we have a sentence given by a respondent and also a guideline sentence, constituting the dataset for the analyses presented in this section.

Considering the means of guideline and respondent sentences it would appear that the American public would like to see convicted felons receive longer sentences than are recommended in the guidelines. As shown in Table 4.1, the mean sentence given by the public was 7.3[6] years in contrast to the average guideline sentence of 5.7. However, differences between guideline and sample medians were not as great, 3 vs. 2.5 years. Using the differences between the averages, it appears that the sample wanted felons to serve 1.6 years longer, but considering the medians, the difference in desired sentence lengths was .5 years. Not surprisingly, the inter-quartile ranges also differ, with the sample sentences spanning a wider distance than the guideline sentences.[7]

In many instances means and medians do not reveal all that should be known about a data set because these measures tell us little about how individual values are distributed. The full distribution of guideline sentences is shown in Figure 4.1 and that of the sample in Figure 4.2. There is an overall resemblance between the two distributions, with sentences in both clustering toward the left sides of the two histograms indicating that most vignettes were given short sentences by respondents and that most crimes were treated that way as well by the guidelines. However, there are also important contrasts: First, the histogram for sample sentences is more irregular, brought about because sentences in multiples of five years were favored over intervening sentences lengths; *i.e.*, respondents tended to favor sentences of 5, 10, 15, 20 (etc.) rather than intermediate lengths. Second, sample sentences tended to be both more lenient and harsher, with more sentences under a year in length and also more sentences at the extreme of 39.2 years.

---

[6] These summary measures differ from those presented in Chapter 3 because in this chapter the upper values of sentences given by sample which were larger than 39 years were changed to 39.2 years to be comparable with the way in which Commission staff treats such sentences numerically. This treatment lowers computed means in this chapter when compared to the treatments in other chapters. It should be noted that median values are not affected.

[7] Guideline sentences were constrained to be identical for identical vignettes whereas respondents sentences were not. Thus the distribution of guideline sentences contains no outlying extreme sentences for any Crime Type or Crime Example.

Table 4.1. Means, Medians and Inter-Quartile Ranges of Guideline and Sample Sentences[a]

Measure	Sentence In Years	
	Guidelines	Sample
Mean	5.7	7.3
Median	2.5	3
Inter-Quartile Range	1.1 to 6.5	.83 to 10
N=	68,712	

[a] All sample sentences greater than 39.2 years were recoded to that value in all tables in this Chapter. Respondent outliers were removed from this table and in other tables in this Chapter.

Figure 4.1. Distribution of Guidelines Sentences
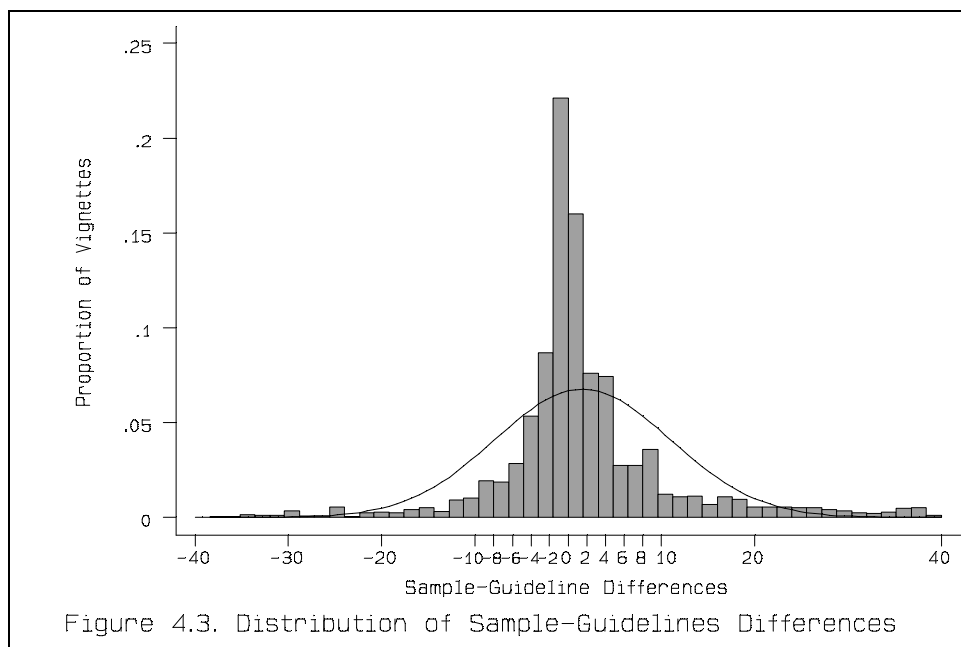
Figure 4.2. Distribution of Sample Sentences
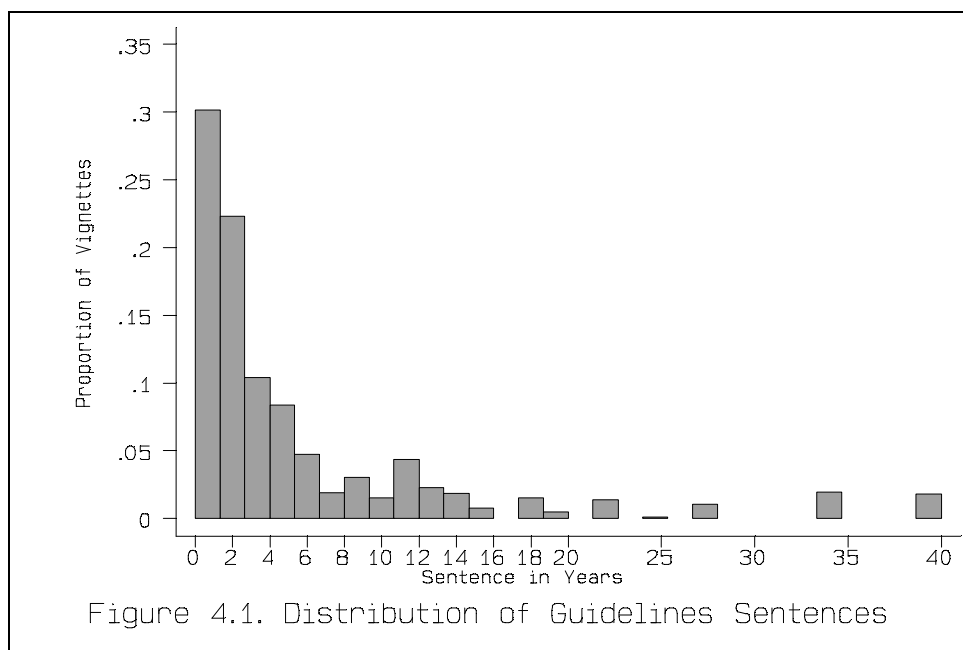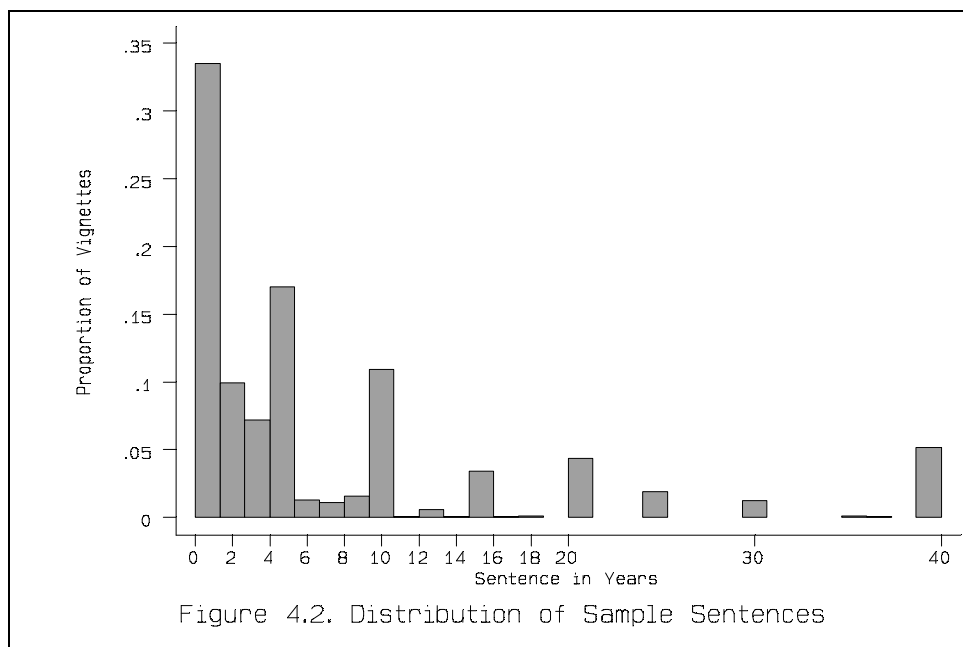
Computing the difference for each vignette between sample and guidelines by subtracting the guideline sentence from the sentence given by the respondent for that vignette, we can examine the distribution of those differences, as displayed in the histogram of Figure 4.3. Note that a positive difference means that the respondent gave a longer sentence than required by the guidelines, and a negative difference means that the respondents gave a shorter sentence than suggested in the guidelines.

In Figure 4.3 a normal curve with the same mean and standard deviation as the distribution of sentencing differences is superimposed on the histogram. The contrast between the normal curve and the histogram shows that the distribution of differences is at the same time more peaked and more dispersed. Half of the differences lie between -1.8 and +3.75, with a median value of 0 and an average of 1.5. In short, there is a fair amount of agreement between respondents and the guidelines, depending on whether discrepancies of that size are regarded as acceptable. About 30 percent of the values are between -1 and +1, indicating that in almost a third of the vignettes the differences between guideline and sample

sentences are a year or less apart. There is also a slight tendency for the sample sentences to be more
lenient, with 59
percent of the
differenc es that
are less than a
y e a r [8] b e i n g
negative.

Figure 4.3. Distribution of Sample-Guidelines Differences

Vignette Level Regression Analysis

How closely do the respondents' sentences correspond to those recommended in the guidelines?
To answer this question in a systematic way, regression analyses were made in which the guideline
sentences were regressed on the respondent sentences for each vignette. In effect, this approach compares
the two measures over the more than 67,000 vignettes.

---

[8] This finding may be an artifact caused by taking the midpoint of the ranges of guideline
sentences. Respondents were allowed to give probation sentences of 0 years, but taking the midpoint of
the most lenient guideline range meant that there were no guideline sentences that were exactly 0.
However, analyses using the lower bound of the guideline sentencing range indicate that only slight
differences occur in most calculations.

Two regression models were employed, with results as shown in Table 4.2. Panel A presents a classic least squares regression equation, and Panel B contains a median regression.[9] Each approach locates a straight line running through points in a plane representing the pairs of guideline and respondent sentences that minimizes the differences between the sentences and that line. The model in Panel A minimizes the squares of those differences and that in Panel B minimizes the absolute values of those differences. In the least squares regression, extreme values strongly influence the location of the line whereas in the median regression extreme values do not play so important a role. Given that respondent sentences do contain many extreme values, the median regression line may represent more closely the general relationship of guideline and respondent sentences.

The interpretation of the coefficients is quite straightforward. In both models, if the guideline and the respondent sentences were identical for each vignette, the value of the regression coefficient would be 1, the intercept term would be 0 and the R^2 value for the equation would be 1. Departures from those values indicate systematic discrepancies. For example, a calculated value of the regression coefficient departing from one means that the two values increase at different rates: for example, a coefficient of 1.5 means that on the average respondents gave an additional 1.5 years for each additional year given by the guidelines. The intercepts indicate the constant value by which guideline sentences differ across the board from respondent sentences: For example, a regression constant of 2.5 indicates that guideline sentences on the average exceed respondent sentences by 2.5 years. The R^2 values are an overall measure of how closely the values cluster around the regression line, one indicating that all the values are on the line and zero indicating that the points are randomly located. The "pseudo R^2" shown for the median regression also has a similar interpretation. In any event, values lower than .5 indicate very modest clustering about the regression line.

The classic regression in Panel A accounts for 22 percent ($R^2 = .22$) of the variance, with an intercept of 2.95 and a regression coefficient of .38. The intercept can be interpreted as the constant difference between sample and guideline sentences and the regression coefficient shows the increment in sample sentences that is associated with each unit change in guideline scores. These results indicate that guideline sentences are systematically higher than respondent sentences by a constant of about three years. However, a one year increment in respondent sentences is only accompanied by .38 years increment in guideline sentences. In other words, the guidelines start off with higher sentences than the respondents and guideline sentences increase by about .38 years for each additional year of respondent sentences. It should also be noted that the least squares regression line does not fit the points very well: R^2 is a very modest .22

The median regression shown in Panel B of Table 4.2., is an equation which uses the medians of the sample sentences and the guideline sentences. Because this approach is based on minimizing the absolute deviations around the regression line, it is much less sensitive to extreme values than the classic regression. The median regression equation has a smaller intercept, 1.29 and a regression coefficient, .30. In other words, the median regression results indicate that the guidelines give higher sentences than respondents, about one year more, and that guideline sentences increase by .3 years for each year increase

---

[9] Median regression is a special case of quantile regression methods. The general approach is recommended when the distribution of values of the dependent variable is skewed. An extended discussion of this approach can be found in Charles F. Manski *Analog Estimation Methods in Econometrics.* Chapman Hall, NY, 1988.

in the respondent sentences. The median regression results do not show the guidelines as being harsher to the same degree as the classic regression results because outliers are not given as much weight. Note also that the predictive power of the median regression is much less than that of the classic regression, the "pseudo-R^2" being .09.

Table 4.2. Regressions of Guideline Sentences on Respondent Sentences. (Units are vignettes.)

A. OLS Regression

$$S_{gi} = 2.95 + .38S_{ij} \quad : \quad R^2 = .22 \qquad N = 67,286$$

B. Median Regression

$$S_{gi} = 1.29 + .30S_{ij} \quad : \text{Pseudo } R^2 = .09 \quad N = 67,286$$

Where: S_{ij} designates sentence given to vignette, i, by respondent, j.

S_{gi} designates Guideline sentence for vignette, i.

All coefficients are significantly different from 0 at better than the .0001 level. In addition, the regression coefficients are also significantly different from 1.00 at better than .0001.

All of the regression results shown in Table 4.2 are highly significant with standard errors that are many magnitudes smaller than the coefficients, about 20 times smaller for the regression constant and 120 times smaller for the regression coefficient. In short, these findings are extremely unlikely to be the result of sampling variation.[10]

---

[10] Because each respondent contributed 42 vignettes, the vignettes are not completely independent. That is, the vignettes rated by a respondent were more likely to be treated somewhat the same way by that respondent and differently from other respondents. As a consequence, standard errors computed in the traditional way tend to be under-estimated.

To gauge the impact of the possible intra-respondent dependence, four random samples of

The regressions in Panel A and B are very much alike, indicating that it makes little difference whether we consider means or medians as measures of central tendencies in either guidelines or respondent sentences. Both regression equations show that on the level of vignettes, the correspondence between sample sentences and guideline sentences is not very high. Although the two kinds of sentences tend to go hand-in-hand, there is also a lot of variation in that overall correspondence, with many respondents giving sentences that are higher and lower than the guidelines.

An important source of the considerable variability of respondent sentences is that they are not very reliable: that is, when the same vignette is rated twice by the same respondent, the sentences given tend to differ. A feature of the vignette design used in this study allows the direct computation of the reliability of respondent sentences. Two vignettes were administered in identical forms to all respondents as "practice" vignettes, as described in Chapter 3. Because the remaining 40 vignettes rated by the respondent were generated by randomly choosing levels from each of the dimensions, some respondents' booklets contained duplicates of the practice vignettes. The practice vignette containing the most serious crime, bank robbery using a gun, was rated twice by 248 respondents.[11] The overall correlation between the two responses was .62, a modest degree of reliability at best.

A reasonable interpretation of the modest degree of reliability in respondent sentencing is that respondents did not distinguish clearly among sentences that are of the same magnitude. A sentence, say, of four years is not seen as very different in severity from a sentence of six years. This interpretation is bolstered by the fact that respondent sentences tended to cluster around sentences that were multiples of five, as shown in Figure 4.1 earlier. Respondents were consistent only in giving the same sentence magnitude and not the same number of years when rating the same vignette twice. In short, at the level of individual respondent sentencing behavior there is evidence that "measurement error" is large enough to account for much of the discrepancies between the guidelines and public opinion.

Taking Response Sets Into Account

---

vignettes were drawn, each consisting of one vignette from each respondent. The least squares regression results follows:

Replicate	Constant	b	R^2	N
1	3.1	.45	.23	1611
2	3.0	.36	.18	1615
3	2.6	.45	.21	1596
4	3.2	.35	.17	1594

In each replicate the standard errors are very small, all coefficients are significant, and, as to be expected not very different from the results in Table 4.2. Whatever intra-respondent consistency may exist cannot account for the highly significant results shown in Table 4.2.

[11] The other practice vignette involved personal use marijuana possession. This is the least serious crime used in the study and hence so little respondent variation (the modal sentence was probation) that reliability was not an issue.

In the analysis of the last section the implicit assumption was made that all respondents had the same understanding of the sentencing metric: That is, a year in prison means the same to one respondent as a year in prison to any other respondent. That assumption may not be justified. Perhaps respondents had different interpretations of prison time, some holding, for example, that a year in prison was a very severe sentence and others holding that a year was not very severe. This might lead some respondents to give longer sentences than others even though all wanted to impose sentences of the same severity. In that event, some portion of the differences between respondent and guideline sentences might well be due to varying "response sets", systematic differences from respondent to respondent in the calibration of sentences. In other words, one respondent may have a general tendency to give long sentences whereas another may be inclined to give generally short sentences.

Although we do not have a very satisfactory way of measuring such response sets, we can take advantage of the "standard vignettes" discussed in Chapter 3, especially the standard vignette concerning bank robbery. Because that vignette was administered to each respondent and hence depicts the same crime and offender, we can use the sentences given to the vignette by each respondent as a proxy for each respondent's response set, reasoning that because the bank robbery vignette was the same for all respondents, the sentences given to that vignette reflect that respondent's response set at least in part.

Table 4.3. Regressions of Sample Sentences on Guideline Sentences Holding Response Set Proxy Constant.

A. OLS Regression

$$S_{gi} = 3.14 + .40S_{ij} - .06V_j : R^2 = .23 \qquad N = 65,749^a$$

B. Median Regression

$$S_{gi} = 1.36 + .29S_{ij} - .02V_j : \text{Pseudo-}R^2 = .10 \qquad N = 65,749^a$$

Where: S_{ij} designates sentence given to a non-standard vignette, i, by respondent j.

S_{gi} designates Guideline sentence given to i.

V_j is the sentence given by respondent, j, to the standard bank robbery vignette

All coefficients are significant at better than the .0001 level.

[a] All standard vignettes involving bank robbery were removed.

Table 4.3 shows how the regression equations presented in Table 4.2 are modified when the sentences given to the standard bank robbery vignette are held constant. Note that the standard vignettes are excluded from the dependent variable. Although the regression coefficient for the guideline sentences is virtually unchanged, varying only in the second decimal place, the regression constants are influenced.

The regression constants are increased slightly. The regression coefficient for the proxy standard vignette is not of much interest in itself although as expected, it is both negative and statistically significant.

The findings of Table 4.3 indicate that respondent response sets play only a minor role in the difference between guideline sentences and those given by respondents, at least to the extent that the sentences given to the standard vignette reflect such phenomena: the resulting R^2s are only slightly higher compared to those shown in Table 4.2. This finding means that the variability from respondent to respondent is not simply a matter of different sentence calibrations. Rather the differences are due to other causes.

Sample-Guideline Rank Order Correspondences for Individual Respondents

An alternative approach to assessing the correspondence between guideline and sample sentences is to examine how closely the rank ordering of sentences given to vignettes by each respondent agrees with the guideline rank ordering. This approach emphasizes the relative ordering of sentences as opposed to their numerical values. A high level of rank order agreement between a respondent and the guidelines means that both agree on which vignettes deserve longer and which deserve shorter sentences although a given respondent may give much longer sentences (or much shorter) sentences overall.

Table 4.4. Distribution of Rank-Order[a] Correlations Between Respondents' and Guideline Sentences.

Spearman Rank Order Coefficients	
Measure	**Value**
Mean	.57
Median	.59
Inter-Quartile Range	.48-.67
Standard Deviation	.15
N=	1628[b]

[a] Spearman rank-order correlations (rho).
[b] Outlier respondents not included.

A rank order correlation was computed for each respondent for the 42 vignettes rated by him or her. The resulting 1,682 coefficients are summarized in Table 4.4. By and large there are considerable

amounts of agreement on the ranking of vignettes: the mean coefficient is .57, quite close to the median, .59. In addition, the coefficients cluster quite closely together: the standard deviation is .15 and the inter-quartile range is .48-.67, indicating that fair agreement is widespread throughout the sample.

Table 4.2 indicates that agreement between guideline and respondent sentence lengths is modest whereas the rank ordering agreement shown in Table 4.4 is fairly high. These two sets of findings are not contradictory: they indicate that the guidelines and respondents agree more on which crimes are serious and which are not but that they agree less on the specific appropriate sanctions to be applied.

Sentences for Crime Types

The vignettes were designed to represent specific concrete cases describing convicted felons and their crimes. At that level, the fit between what the respondents want as sentences and what the guidelines prescribe is not very close. Although the two tend to go hand in hand, there were differences of kind and degree between the two. However, it can be argued that it is not necessary to have close agreement about specific cases as long as there is fairly close agreement among the central tendencies of sentencing about classes of crimes.

Table 4.5 lists the guideline and respondent sentences for the 20 different Crime Types included in the study. To simplify presentation, the Crime Types are ordered by the mean sentence given under the guidelines. In the first column, one can see that kidnapping is the most heavily penalized crime in the guideline schedule (25.7 years), drug trafficking is the second most heavily penalized crime (12.5 years), and so on. The second column shows the mean sentences given by respondents. For example, the mean respondent sentence for kidnapping is 23 years, the mean respondent sentence for drug trafficking is 13.5 years, and so on. Columns 3 and 4 show the median sentences given under the guidelines and by respondents respectively.

Respondent sentence measures in Table 4.5 and in Table 4.7 differ from values shown in Chapter 3 because of the truncation of extreme sentences described earlier.

Note that respondent sentences have means and medians which are generally far apart. We favor using the medians of respondent sentences for further analyses because they will be unaffected by the arbitrary coding of life and death sentences (which fall at the tails of the sentencing distribution.) In contrast, the medians and the means for guideline sentences are much closer together, a consequence of the fact that the guideline sentences as calculated for some specific crime were constrained by the guideline range. Accordingly, guideline sentences are quite well represented by their means.[12]

Finally, the next two columns show the interquartile ranges for the guideline sentences and the respondent sentences. As shown in the inter-quartile ranges, by and large, there is more variability in the respondent sentences, which is not surprising. Variation in the guideline sentences is solely a function of vignette characteristics, while variation in respondent sentences is a function of vignette characteristics *and*

---

[12] Because guideline sentences are fixed for all identical vignettes, there are no outliers in these sentences. Hence the mean is a good descriptive measure of their central tendencies. In addition, the findings in the analyses of the next few pages are not strongly affected by whether means or medians are used to represent guideline sentences.

respondents' characteristics. The only exception to the general pattern is for drug trafficking where the inter-quartile range for respondent sentences is a bit smaller. The less variable respondent sentences reflect the fact that respondents did not treat the four kinds of drugs as differently as the guidelines. In particular, respondents did not regard trafficking in crack cocaine as being especially heinous compared to trafficking in powder cocaine, heroin or even marijuana whereas in the guidelines crack trafficking is punished more severely.

Comparing the two series, it is clear that sometimes guideline sentences are more lengthy and sometimes respondent sentences are more lengthy. For example, the guideline sentence is longer for drug trafficking, and bank robbery, but shorter for extortion and forgery.

Table 4.5 Sentences for Crime Types: Guideline and Sample Sentences Compared. Crime Types Arranged in Descending Order of Average Guideline Sentences

Crime Type	Means (Years)		Medians (Years)		Inter-Quartile Range		N of Vignettes
	Guide-lines	Sample	Guide-lines	Sample	Guide-lines	Sample	
Kidnapping	25.7	23.0	39.2	25	27.8	34.2	1,396
Drug Trafficking	12.5	13.5	7.3	10	19.5	15	13,413
Bank Robbery	10.9	9.6	11.3	5	4.5	7	5,499
Street Robbery	9.7	9.2	9.1	5	6.8	7	1,328
Food & Drug	7.3	13.3	6.5	9	10.8	18	1,989
Extortion	3.8	8.6	3.1	5	4.8	8	683
Money Laundering	3.8	5.0	3.1	3	4.1	4.2	1,995
Major Fraud	3.5	6.3	3.5	3	3	9.2	4,035
Civil Rights	3.1	3.1	4.3	1	3.2	3.6	3,343
Environment	2.9	3.4	2	1	3.4	5	3,315
Firearms	2.5	4.8	2.5	2	2	4.5	3,370
Tax	2.5	4.4	2.2	2	1.8	4.5	4,014
Forgery/Counterfeit	2.4	6.3	1.8	5	2.3	8	2,049

Antitrust	2.3	4.7	2	2	.5	4.5	1,353
Bribery	2.1	3.0	2.5	1	1.4	3.5	2,597
Minor Fraud	1.8	4.4	1.8	2	1.3	4.2	2,709
Larceny	1.7	4.9	1.2	3	1.5	4.1	1,908
Embezzlement	1.3	4.7	.9	2	1.7	4.2	1,945
Immigration	.8	4.3	.8	2	.58	4.5	3,333
Drug Possession	.6	2.0	.25	.5	.5	2	7,012

Figure 4.4 shows the same data in a scatterplot format. Median respondent sentences are represented on the horizontal axis and mean guideline sentences are represented on the vertical axis. The diagonal line is the least squares line resulting when mean guideline sentences are regressed on median respondent sentences. The two lines banding the regression line indicate where guideline sentences would fall two standard errors above and below the predicted regression values. Points lying outside the bands are identifiable as statistical outliers which are distant from the regression line far beyond chance expectations.

Crimes with deviations lying outside the "confidence band" are labeled. No crimes fall below the confidence band and only two Crime Types lie above the band: bank robbery and street robbery, both Crime Types for which guideline sentences are much higher than respondent sentences.

The first impression is that there is, on the average, remarkable comparability between mean guideline sentences and the median sentences desired by the public for Crime Types. A more precise evaluation can be obtained by testing the null hypothesis that the intercept of the regression line is 0.0 and the slope of the regression line is 1.0. If the intercept is 0.0, there is no systematic tendency over the full set of Crimes Types for mean guideline sentences to be longer or shorter than median respondent sentences. If the slope is 1.0, a one year increment in median respondent sentence length is associated on the average with a one year increment in mean guideline sentence; on the average, the two sentencing distributions move in step with one another over the full set of Crime Types and in that sense, the *structure* of the sentencing is the same. The estimated regression equation has a regression coefficient 1.02 and an intercept is .56, neither statistically different, respectively, from 1 and 0.0. The departures are small in practical terms. Stated differently, given a respondent median sentence for a Crime Type, the best predicted guideline sentence is the actual calculated mean guideline sentence.

The fit is also impressive, with an R^2 for the equation of .87. A key implication is that when there are departures from the regression line, they tend to be small. At the Crime Type level of aggregation, one can "explain" most of the variation in median guideline sentences from information about median respondent sentences.

Because kidnapping, the upper right hand point of Figure 4.4, receives sentences that are much longer than any of the other Crime Types, one could suspect that much of the overall correspondence is dominated by the Crime Type of kidnapping. However kidnapping is not particularly influential. The least

squares regression lines with and without kidnapping are virtually identical: Without kidnapping the intercept is .44 and the slope is 1.05, whereas the regression line with kidnapping included has an intercept of .56 and slope of 1.02.[13]

Perhaps the major lesson from Figure 4.4 is that there is fairly strong agreement on aggregated sentences for Crime Types. However, as a unit, Crime Types are coarse; overall conclusions heavily depend on how the vignettes were designed and then aggregated. For example, we suspect that if the design had allowed for the killing of a teller or witness in some of the bank robbery vignettes, the median sentences for bank robbery might have looked a lot like the median sentences for kidnapping. But most important, in the real world, sentences are given out for specific offenses and not for the broad categories represented by Crime Types. These considerations argue for looking at the correspondence between guideline and respondent sentences using more specific offenses.

---

[13] The adjusted R^2 for the equation including kidnapping is .87 whereas that for the equation excluding that crime is .60. Although kidnapping does not meaningfully influence the location of the least squares line, it adds dramatically to the explained variance.

Figure 4.4. Crime Types Scatterplot
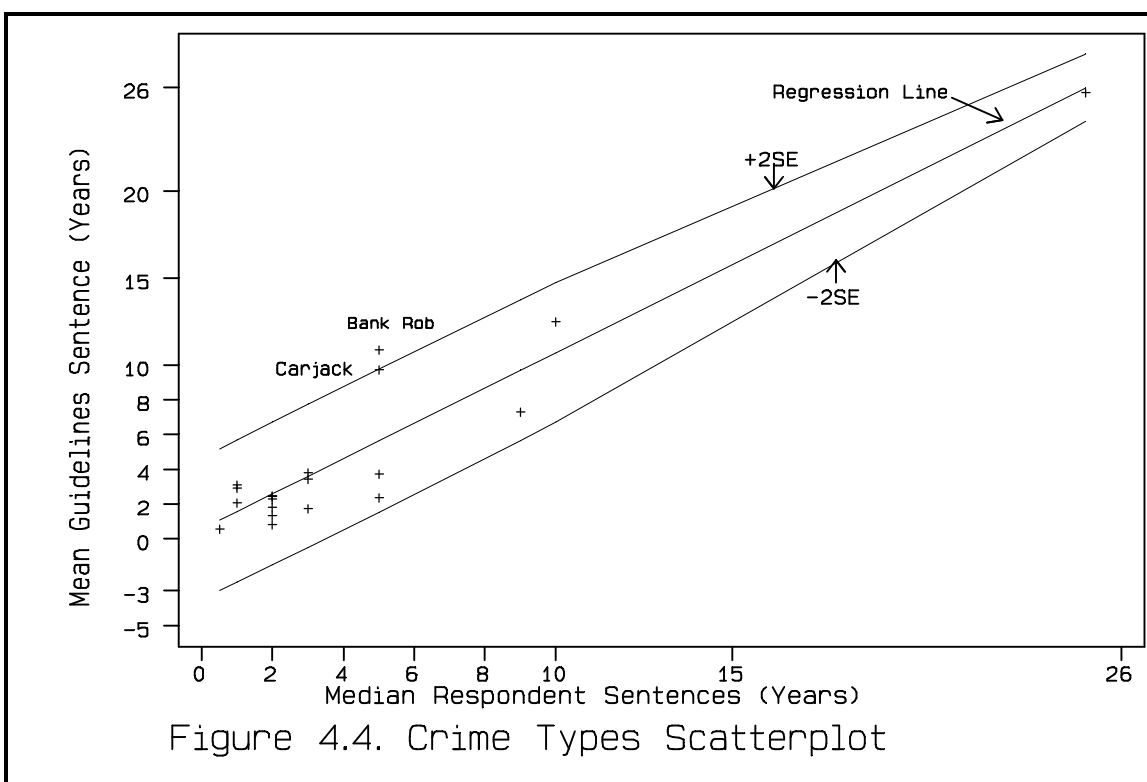
The 20 Crime Types are each composed of several examples. Each example is a description of a concrete instance of one of the Crime Types. (The Crime Example descriptions are presented in detail in Figure 3.2.) A strong argument can be made that it is the degree of correspondence concerning examples that should be a matter of concern, because the examples are closer to the way in which crimes are perceived by the public. Accordingly, a finer-grained approach to comparing guideline and sample sentences is to examine that correspondence at the level of the 73 Crime Examples used.

Table 4.7 shows mean and median sentences under the guidelines and given by respondents for the 73 Crime Examples. Overall, going to the more specific Crime Examples produces greater heterogeneity in both guideline and respondents sentences. In addition, there are more and sometimes greater disparities between the mean guideline and the median respondent sentences.

In Table 4.7, by far the largest disparity between the mean guideline sentences and the median respondent sentences is for trafficking in crack cocaine; the guideline mean sentence is almost 12 years longer than the respondent median sentence (21.8 years versus 10 years). Interestingly, the disparity for trafficking in marijuana is also large, but in the opposite direction (4.4 years versus 2.5 years). The guidelines make an decided distinction between trafficking in crack cocaine compared to marijuana. The public agrees that trafficking in crack is more serious than trafficking in marijuana, but only a little bit more serious.

Table 4.7. Guideline and Sample Sentences Compared for 73 Crime Examples. Crime Examples Ranked in Descending Order by Mean Guideline Sentences

Crime Example	Sentences Given (Years)				N of vignettes Rated
	Guideline Mean	Sample Mean	Guideline Median	Sample Median	
Kidnapping: Victim killed	39.1	33.9	39.2	39.2	724
Drug Trafficking: Crack	21.8	14.2	22	10	3,281
Bank Robbery: Weapon used with major injury to victim	17.5	17.9	17.6	15	635
Bank Robbery: Weapon used with minor injury to victim	14.1	14.4	14.1	10	616
Drug Trafficking: Cocaine	12.1	14.2	9.1	10	3,387
Drug Trafficking: Heroin	11.8	14.0	9.1	10	3,359
Kidnapping: No harm to victim	11.5	11.3	11.3	6	672
Robbery: Carjacking	11.4	9.8	11.3	5	732
Bank Robbery: Weapon fired with no harm to any victim	11.3	7.8	11.3	5	2,240
Food & Drug: Poisoning over-the-counter drugs	10.5	19.7	9.4	15	681
Bank Robbery: Weapon used but not fired	9.24	8.1	8.1	5	681
Environment: Plant discharging toxic wastes into stream	8.6	4.0	8.1	2	664
Robbery: Convenience store	7.6	8.4	6.5	5	596
Money Laundering: coin dealer arranging false sales to criminals	7.3	6.7	7.3	4	669
Bank Robbery: Bomb threat used	6.9	8.7	6.5	5	666
Food and Drug: Drug manufacturer concealing bad side effects	5.9	11.4	1.8	5	631
Bank Robbery: No weapon used	5.6	5.6	4.8	4	661
Food & Drug: Marketing drugs after false testing	5.4	8.5	1.8	5	677

Crime Example	Sentences Given (Years)				N of vignettes Rated
	Guideline Mean	Sample Mean	Guideline Median	Sample Median	
Fraud: Bank officer causing S&L failure	5.4	6.1	4.8	3	660
Forgery: Counterfeiting currency	4.7	7.6	3.8	5	704
Drug Trafficking: Marijuana	4.4	11.6	2.5	8	3,386
Civil Rights: Police beating unresisting motorist	4.3	4.2	4.3	2	675
Civil Rights: Police beating minority motorist	4.3	3.7	4.3	2	709
Civil Rights: Police beating motorist resisting arrest	4.3	1.9	4.3	0.5	636
Extortion/Blackmail	3.8	8.6	3.1	5	683
Fraud: Selling defective helicopter parts to government	3.7	11.2	3.8	10	672
Firearms: Possession of sawed-off shotgun	3.4	4.4	3.1	2	670
Bribery: Bribing local official	3.4	3.5	3.1	2	686
Firearms: Dealer selling firearms to felons	3.0	7.1	2.8	5	657
Fraud: False mortgage application with intent to pay mortgage	3.0	2.0	3.1	.5	678
Fraud: Selling worthless stocks and bonds	2.9	7.3	3.1	5	680
Fraud: Doctor filing false Medicare claims	2.9	6.9	3.1	5	684
Fraud: Company official making use of inside information	2.9	4.3	3.1	2	661
Environment: Plant discharging hot water into stream	2.8	2.2	2.5	0.7	658
Firearms: Dealer failing to keep proper sales records	2.6	5.8	2.5	3	725
Tax: Failure to file tax returns	2.5	4.4	2.5	2	1,341
Tax: Under-reporting income on tax returns	2.5	4.4	2.5	2	1,340

Crime Example	Sentences Given (Years)				N of vignettes Rated
	Guideline Mean	Sample Mean	Guideline Median	Sample Median	
Firearms: Felon owning handgun	2.5	3.7	2	2	691
Bribery: Local official taking bribe	2.5	2.6	2.5	1	615
Money Laundering: Coin dealer failure to file required forms	2.4	3.8	2.2	2	660
Anti-Trust: Bid rigging on contracts	2.3	5.9	2	4	642
Larceny: Buying and selling stolen goods	2.3	4.5	2	2	619
Tax: Promoting illegal tax shelter	2.3	4.3	1.8	2	1,333
Anti-Trust: Price fixing	2.3	3.6	2	1	711
Fraud: Soliciting funds for non-existing charity	2.2	4.9	2	3	661
Environment: Factory failure to install smoke pollution devices	2.2	3.8	2	1	674
Fraud: False Mortgage application with no intent to pay back mortgage	1.8	3.8	1.8	2	628
Embezzlement: Postal worker taking postal funds	1.7	5.6	1.8	3	672
Larceny: Stealing mail	1.7	5.5	1.2	3	636
Money Laundering: Bank official failing to file proper forms	1.7	4.5	1.8	2	666
Fraud: Using stolen credit cards	1.7	4.2	1.5	2	701
Fraud: Writing bad checks	1.6	4.6	1.5	2	719
Embezzlement: Bank officer stealing bank funds	1.3	4.3	0.9	2	663
Civil Rights: Harassing new neighbor to get them to move out	1.3	3.2	1.1	1	650
Civil Rights: Vandalizing place of worship	1.3	2.5	1.1	1	673
Forgery: Using stolen credit card	1.2	6.1	0.9	4	664
Forgery: Writing bad checks	1.2	5.0	1.2	3	681

Crime Example	Sentences Given (Years)				N of vignettes Rated
	Guideline Mean	Sample Mean	Guideline Median	Sample Median	
Larceny: Stealing Property	1.2	4.8	0.8	3	653
Bribery: Bribing private company agent to give contract	1.2	3.2	1.1	2	657
Immigration: Smuggling aliens endangering their safety	1.1	7.1	0.9	5	727
Immigration: Smuggling aliens for profit	1.1	5.8	0.9	4	649
Bribery: Government agent taking bribe	1.1	2.6	0.8	1	639
Embezzlement: Bank employee taking bank funds	0.9	4.3	0.6	2	610
Drug Possession: Crack	0.8	2.9	0.8	1	1,361
Immigration: Illegal re-entry into US	0.8	2.7	0.8	0.9	631
Drug Possession: Heroin	0.8	2.6	0.8	0.9	1,368
Firearms: Felon owning hunting rifle	0.7	3.0	1	1	627
Immigration: Smuggling alien family members into US	0.6	3.4	0.4	1	668
Immigration: Illegal entry into US	0.6	2.4	0.4	0.7	658
Environment: Killing bald eagle, an endangered species	0.6	2.4	0.4	0.9	669
Environment: Illegal logging on federal lands	0.5	4.6	0.3	2	650
Drug Possession: Cocaine	0.5	2.8	0.4	0.9	1,329
Drug Possession: Marijuana	0.3	1.0	.2	0	2,954

There are additional Crime Examples in which the guideline means are noticeably larger than respondent means, including carjacking, bank robbery, several environmental crimes, and one of the civil rights crimes. There are also examples in which the public apparently wants more severe sentences than the guidelines, including several immigration crimes, the sale of defective helicopter parts to the federal government. Overall agreement is clearly tempered by strong differences surrounding specific offenses.

Figure 4.5 graphs the relationship between the two series. The two lines above and below the regression line form a "confidence" band. Falling far above and outside the band is the Crime Example trafficking in crack cocaine, clearly representing the greatest disparity between the guidelines and the respondents. Also above the confidence band were the Crime Examples involving a bank robbery in which a gun was fired but no one was injured; carjacking in which the victim is not injured; and the environmental violation in which toxic waster were released into a stream.

Crimes falling at or below the lower line of the confidence band are Crime Examples in which respondents gave much longer sentences than the guidelines. Only two Crime Examples are so identified: a fraud crime in which defective helicopter parts were sold to the government; and adding poison to over the counter drugs. Significantly, both these Crime Examples involve potential physical danger to persons.
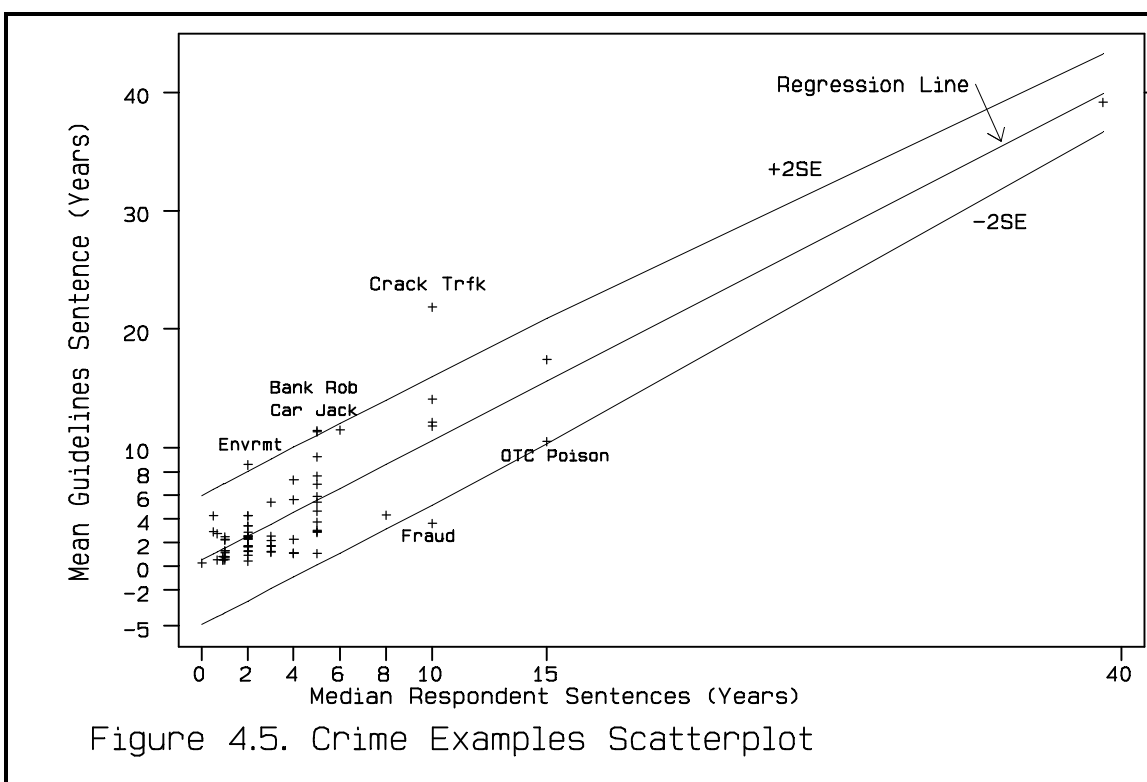
Figure 4.5. Crime Examples Scatterplot

ween mean guideline sentences and median respondent sentences. The regression line has an intercept of .55 and a slope of 1.00, values that are not statistically different from 0.0 and 1.0 respectively.[14] In addition, the adjusted R^2 for the regression is .78;[15] the few striking departures from the regression line still left a very good fit intact. In short, on the one hand, we find once again that overall the guidelines map well onto the public's views. On the other hand, the few departures seem all the more anomalous by contrast.

Combining Prior Record and Crime Examples

At the very end of the process of determining guideline sentences, each crime is combined with the convicted person's prior record of felony convictions. The design of the vignettes also followed that process with each vignette randomly assigned a prior record of either none, two, or four felony convictions. This provides an opportunity for a much finer aggregation by combining Crime Examples and prior record.

---

[14] The regression results using the median guideline sentences were very close; the intercept was .23 and the slope .98, not statistically different from 0.0 and 1.0 respectively.

[15] We also tested for regression line differences when the Crime Example of kidnapping and killing the victim was omitted. The resulting regression equation was virtually identical to the line computed when that Crime Example was included, although the adjusted R^2 for that line dropped to .58.

There are 175 such combinations,[16] too many to show in a table. However, a scatterplot can be more easily comprehended: it is shown as Figure 4.6.

Figure 4.6 has the same structure as shown in prior figures. It plots the mean guideline sentences and the median respondent sentences for each of the 175 combinations. Three lines are shown, the upper representing sentences two standard errors above the predicted guideline sentence, the middle line being the regression line, and the lower line representing two standard errors below the regression line. Points at or outside the confidence band are identified and named, the number in parentheses representing respectively a prior record of four previous convictions (the number 3); a prior record of two previous convictions (the number 2); and no prior record (the number 1).

The least squares regression line has an intercept of .45 and a slope of 1.02. As before, we cannot reject the hypothesis that the intercept was 0.0 and the slope 1.0. Once again, the structure of the two sentencing distributions is, on the average, the same. The adjusted R^2 for the regression is .80, about the same as for the Crime Examples. In short, taking the effects of prior record into account, the correspondence between guideline and respondent sentences remains high. This finding also indicates that the respondents and the guidelines agree on prior record adjustments to sentences.

All that said, there are 12 combinations that lie outside the confidence band, nine of which are offenses to which the guidelines give much higher sentences than the respondents. Conspicuous among the outliers are the three crack trafficking offenses: all of the combinations involving that crime. Clearly the guideline sentences for this crime are much harsher than respondents desired and even more so for offenders with prior records. The other combinations in which guidelines are harsher than respondents include: carjacking committed by an offender with four prior convictions but with no resulting physical harm to the victim, bank robbery in which the robber has four prior convictions, in one case brandishing a gun and in the other case discharging the gun harmlessly into the bank ceiling, and the environmental crime of discharging toxic chemicals into a stream (all levels of prior convictions).

There are three combinations in which respondents give much longer sentences; two involving felons with either two or four prior felonies putting poison into over-the-counter drugs, and the other a fraud in which a manufacturer sold defective helicopter parts to the government.[17] In the latter example, respondents were likely reacting to the potential for physical harm to helicopter passengers.

---

[16]For some of the Crime Examples involving persons in occupations denied to those who have prior felony convictions, e.g. bank officials, elected officials, or police, vignettes were given the no prior record description. Hence only one combination was possible for such vignettes.

[17]This Crime Example was one in which no prior record was given to all vignettes with this example.

Figure 4.6. Combined Examples And Prior Record Scatterplot
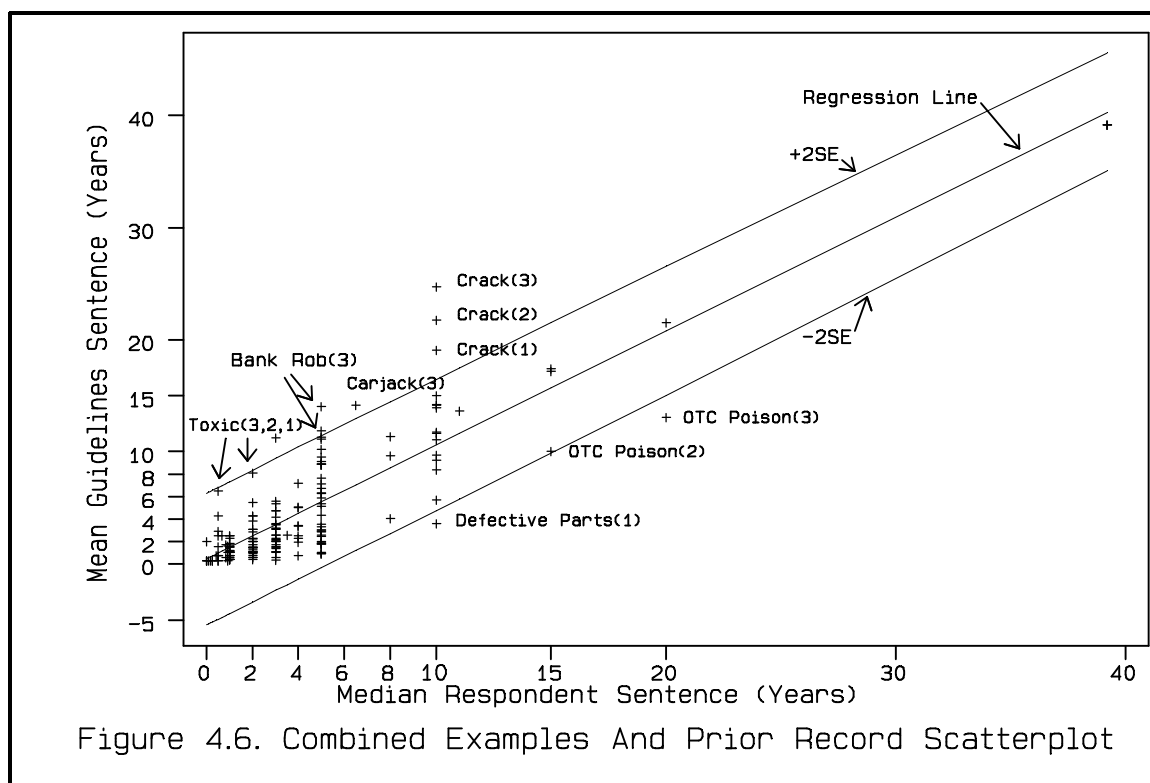
Overall, median respondent sentences were close matches to the mean guideline sentences suggesting that the public agrees with the guidelines about the weights to be given to prior criminal records as well as the punishments to be meted out for crimes.

Agreement Over Crime Dimensions

For each of almost all of the Crime Types and Crime Examples, the vignettes incorporated variations in features of the crime that were recognized in the guidelines as justifying either an enhancement or diminution of the sentences for the crime in question. Many of these guideline sentence modifications were represented in the vignettes as levels of a dimension. (Figure 3.1 contains the texts of all the dimensions and levels used.) Accordingly, the crime of bank robbery was shown in vignettes as varying in the dollar amounts of the robberies as well as in the extent to which handguns were used in the commission of the crimes. One guideline sentence modification applies to all the crimes in which an offender could have a prior record: Sentences are to be increased according to the previous criminal record of the offender. We also added dimensions to the vignettes, such as the gender of the criminal, which are not used in the guidelines as bases for the modification of sentences.

The dimensions are another potential source of differences between guideline and sample sentences. For example, it may be that respondents weighted previous imprisonment records differently

than the guidelines, giving larger (or smaller) sentence enhancements for the number of previous felony convictions.

Note that later chapters (Chapter 5 through 9) provide details on how the sample used these dimensions in arriving at sentences. In this section, attention is focussed on comparability with guidelines.

Table 4.8. Guideline-Sample Comparisons in Dimension Weighting

A. Significant[a] Dimension Weighting Differences Between Guideline and Sample Sentences

Dimension Involved	Maximum Years Difference[b]
Drug Trafficking: Drug Type	14
Drug Trafficking: Drug Amount	15
Drug Trafficking: Role in Transaction	6
Drug Trafficking: Weapon Use	5
Minor Fraud: Dollar Amount	2
Firearms: Number of Weapons	2
Firearms: Knowledge of Criminal Use	3
Larceny: Dollar Amount	2
Bank Robbery: Dollar Amount	2
Street Robbery: Convenience Store; Weapon Use	2
Embezzlement: Dollar Amount	3
Embezzlement: Degree of Planning	2
Food and Drug: Harm to people	3
+Environment: Logging; Habitat Destroyed	1
+Environment: Logging; Streams Polluted	1
+Environment: Air Pollution: Foul Smells	1
+Environment: Waste Water; Fish Kills	1
Tax: Dollar Amount	1
Tax: Criminal Source of Income	2
Bribery: Dollar Amount	1
Drug Possession: Drug Type	1
Forgery: Dollar Amount	2
Money Laundering: Knowledge of Criminal Source	1
*Gender of Offender	.5
*Family Status of Offender	.4

B. No Significant[a] Differences in Dimension Weighting Between Guideline and Sample Sentences

Previous Imprisonment Record	0
Major Fraud: Dollar Amount	1
Larceny: Degree of Planning	1
Street Robbery: Convenience Store; Harm	2
Street Robbery: Convenience Store: Money Amount	2
Street Robbery: Carjacking: Money Amount	0

*Civil Rights: Minority Status of Motorist	0
*Civil Rights: Denomination of House of Worship	0
Antitrust: Dollar Amount	0
Antitrust: Role in Violation	0
*Food & Drug: Drug poisoning:	
Amount of Company Money Lost	0
+Environment: Logging; Damage to Aesthetics	0
+Environment: Logging: Damage to Watershed	0
+Environment: Air Pollution: Paint Damage	0
+Environment: Air Pollution: Tree Damage	0
+Environment: Air Pollution: Respiratory Damage	0
Extortion: Dollar Amount	1
Money Laundering: Dollar Amount	1
*Employment Status of Offender	0

[a] Significance was judged taking the N of vignettes into account: If vignette N was greater than 10,000, significance meant that p< .0001. If vignette N was less than 10,000 and more than 999, p< .001. If vignette N was less than 1,000, p < .01.

[b] Maximum difference was defined as the range of departure from guidelines. For example, if for one drug type dimension, the sample sentences for one level is ten years less than guidelines and for another five years more than guidelines, the maximum was calculated as 15 years.

* Indicates dimension that is not recognized in the guidelines as affecting recommended sentences.

+ Indicates dimension which could not be given specific guideline sentence enhancements although recognized in the guidelines as grounds for doing so.

      As discussed earlier in this chapter, aside from Crime Types and prior record, the mapping of guideline sentence enhancements on the vignettes dimensions was not accomplished successfully in all cases. This means that for some of the dimensions, the comparison is between a poorly translated guideline enhancement and the respondents' sentences enhancement for that dimension. Such shaky comparisons will be indicated in the analyses.

      To investigate this issue, we tested whether the differences between guideline and sample sentences across the levels of each dimension were statistically significant.[18] Most of the dimensions were used only

---

      [18] These tests were conducted by calculating the difference between the sample sentence and the guideline sentence given to each vignette and conducting an analysis of variance test of those differences across the levels of each dimension. If the differences between the guideline and sample sentences were not statistically different, say between offenders with different previous felony convictions, then we concluded that the guidelines and the sample did not differ in the weighting scheme each used.   When such differences were found to be statistically different, we concluded that the sample and the guidelines gave different weights to the levels of the dimension in question.

      Because the number of vignettes involved varied from dimension to dimension, we applied

in the Crime Type or Crime Example for which they were relevant. Only four were included in all the vignettes: previous criminal record, gender, family status, and employment status.

All told, there were 44 dimensions used in constructing the vignettes of which 28 were recognized in the guidelines as grounds for specific sentence modifications. Dimensions that are not given specific enhancements in the guidelines are marked with asterisks. Dimensions concerning which there were uncertainties about how to translate the guidelines are marked with +.

In addition, we calculated for each dimension the maximum years the sample sentences departed from the guidelines, the resulting maxima being shown for each dimension.

The outcomes of the statistical tests are shown in Table 4.8. In Panel A are listed the dimensions over which there were statistically significant disagreements between the guideline and sample sentences. In Panel B, the dimensions in which there were no significant disagreements are listed.

The findings in Table 4.8 can be summarized as follows:

1. There were more disagreements than agreements, 25 versus 19, indicating that respondents tended to give weights to most dimensions that were different from those called for in the guidelines.

2. For most dimensions, the maximum sentencing disparities for a dimension were small — typically one or two years.

3. The dimensions used in drug trafficking crimes stand out as producing very large disparities. Respondents did not distinguish as strongly among specific drugs, among trafficking in various amounts of drugs, gave different weightings to the roles played by defendants and the uses of weapons in those crimes.

4. Previous imprisonment records of felons were treated by respondents in the same way as the guidelines, a critical agreement because of the importance given to this dimension in the guidelines. Agreement between the guidelines and respondents of the treatment of this dimension undoubtedly is a strong contributor to the overall close correspondence between the two kinds of sentences.

5. Dollar loss sentence enhancements in the guidelines were a major source of disagreement. Of the 12 such guideline provisions, eight produced disagreements, respondents generally giving smaller additional punishment than the guidelines for increases in the dollar amount losses.

6. The consequences of violations of environmental laws, as described in the vignettes, affected respondent sentencing in four of the nine dimensions whereas the guidelines provided no specific enhancements but permitted judicial departure from the guideline range.

---

different criteria for judging significance as shown in the notes attached to Table 4.8.

7. The gender and family status of the felons depicted in vignettes are not recognized in the guidelines as relevant to sentencing but respondents were more lenient to women and felons with dependents. These differences were very small.

The largest departures from guideline sentence modifications were recorded for the Crime Type of Drug Trafficking. The sample responded quite differently to the type and amount of drugs involved, to the role played by the offender in the drug transaction, and to the use of weapons in drug trafficking. Throughout this chapter it was found that the guidelines treated drug trafficking in the crack form of cocaine much more harshly than the respondents, a finding that shows up in this analysis as a maximum of 14 years disparity in sentencing for drug types. The guideline sentences are also much more sensitive to the amount of drugs involved in the transaction, leading to a maximum disparity in sentencing of about 15 years. The other drug trafficking dimensions also led to large disparities between the sample and the guidelines. Indeed, all the drug trafficking dimensions produced maximum disparities that were greater than for any of the other dimensions.

The importance of these disparities in drug trafficking vignettes in the overall comparison between guideline and sample sentences is shown dramatically if we compute the regression between the two omitting the drug trafficking crimes: As shown in Table 4.2, the R^2 over all vignettes is .22. However, when we omit the drug trafficking vignettes the R^2 rise to .28, a 33 percent increase. The regression coefficients remain essentially unchanged. In short, putting drug trafficking crimes aside increases considerably the vignette level comparability between guideline and respondent sentences.

Summary and Conclusion

We return to the central question posed in the beginning of this chapter. To what extent do the Sentencing Commission's guidelines correspond to the public's views on the sentencing of persons convicted of violating the federal penal code? The findings presented indicate that the guidelines map rather well onto the central tendencies of public wishes. In each instance when we examined aggregate respondent data, we were unable to reject the null hypothesis that the two sentencing distributions had the same structure. Furthermore, because the federal criminal courts deal with crimes of quite a different mix compared to the state courts, which are more likely to receive attention from the mass media, one might expect that the federal crimes might be less well known and hence views on sentencing of such crimes might not be very well structured. In this light, the similarities between the guidelines and public views are remarkable.

With the major exception of drug trafficking crimes, it was also found that the factors considered by the guidelines as grounds for specific kinds of sentence enhancements were also regarded as grounds for such changes in sentencing by the public. Especially important was the close agreement over sentence enhancements justified by the prior criminal records of the offender.

Yet, within the remarkable correspondence between the guideline sentences and respondent sentences, there were several striking disagreements. In particular, the guidelines and the public differ strongly on trafficking in illegal drugs. The guidelines favored very severe punishments for dealing in crack cocaine while the public does not regard trafficking in that drug as more serious than dealing in either powder cocaine or heroin. At the same time, the public desired somewhat longer sentences for trafficking in marijuana. The message from the public may be that trafficking in any illegal substance is major felony, but trafficking in crack cocaine should not be singled out for especially severe punishments.

Less striking, but still important, is that the public appears to regard crimes that endanger the physical safety of victims and bystanders as more serious than the guidelines, as, for example, the Crime Example of adding poison to over-the-counter drugs. In general, concerns about the safety of individuals drive a lot of the variation in sentences for both the guidelines and the public, but these concerns translate more readily into longer sentences when the public's preferences are elicited.

Finally, the guidelines treat more harshly than the public: environmental crimes, violation of civil rights, and certain bribery and extortion crimes. It is difficult to see what these crimes have in common, and we fail to find any general lessons.

At a more general level, our data are consistent with a view that the guidelines are substantially structured by the central tendencies of public opinion. At the same time, there is significant variation around these central tendencies so that particular individuals are unlikely to see in the guidelines a fully accurate reflection of their own preferences; the structure of correspondence only becomes apparent in the aggregate. Thus, there is ample room for visible and heated criticisms of the sentences prescribed by the guidelines even though the guidelines map rather well onto what the public "wants" on the average.

An interpretation of the differences among levels of aggregation is that there is considerable "error" in individual respondent sentences. There is apparently no clear view of an absolute scale of severity of punishment that corresponds directly to lengths of prison sentences. One person's two year sentence may be the equivalent of another's four year sentence. In addition, the differences between sentences are not distinct; respondents who gave a four year sentence on one occasion to a specific crime may give a different sentence on another occasion to the same crime. In other words, the punishment norms of our society are only dimly apprehended by respondents. These "errors" tend to cancel out when responses are aggregated as shown in our analyses at the Crime Type and example levels of aggregation. In line with this interpretation is the finding that respondents with more formal education have smaller "errors" in their sentences than those with less education.[19]

---

[19] Quantile regressions of guideline sentences on respondents sentences for those with college degrees produced a "pseudo R^2" more than twice as large as that for respondents who never completed high school.