

Sentencing Councils: A Study of Sentence Disparity and its Reduction*

Shari Seidman Diamond† and Hans Zeisel††

The fundamental premise in the idea of impartial judges and rules of law is that certain kinds of decision-making, for example, by judges, can by institutional arrangements and role discipline be made to show less variance and less correlation to personal factors than other kinds of decision-making Beyond this there is the question of whether different arrangements will produce different degrees of impersonality in decision-making.¹

HARRY KALVEN, JR.

The severity of a criminal sentence depends to some extent on the identity of the sentencing judge. Comparable cases will not always receive comparable sentences. Although all would acknowledge that such sentence disparity among judges exists, there is disagreement both about the magnitude of the disparity and how it might be reduced.

One institutional arrangement used in some federal courts to alleviate sentence disparity is the sentencing council. The council enables the sentencing judge, before imposing sentence, to meet with his colleagues in order to learn what sentences they would impose if they were the sentencing judge. This article first reviews the problems sentence disparity has created and recounts earlier

† Assistant Professor of Criminal Justice and Psychology, University of Illinois, Chicago.

†† Professor of Law and Sociology Emeritus, The University of Chicago.

* This study is part of a larger research program on the problems of reducing sentence disparity. The research is supported by a grant from the Law Enforcement Assistance Administration to the Center for Studies in Criminal Justice at The University of Chicago Law School. We are grateful to Edwin A. Robson, Chief Judge of the United States District Court for the Northern District of Illinois, and to Judge Hubert L. Will, chairman of the sentencing council in that court; to Jacob Mishler, Chief Judge of the United States District Court for the Eastern District of New York; and to all the judges of both courts for their generous cooperation. The central retention of the sentencing court records by the Probation Department of the Chicago court greatly facilitated our work. Not the least of our thanks go to Judith Peyton and the Computation Center of the University of Illinois in Chicago for their dependable assistance. We also would like to thank our colleague Franklin B. Zimring, the Director of the Center for Studies in Criminal Justice, for his counsel and gentle impatience.

¹ Kalven, *Toward a Science of Impartial Judicial Behavior*, *Symposium: Empirical Approaches to Judicial Behavior*, 42 U. CIN. L. REV. 594 (1973).

efforts to study the phenomenon. It then develops a measure of disparity that might prove useful in future research and uses it to assess the magnitude of sentence disparity in two federal district courts: the Northern District of Illinois (Chicago) and the Eastern District of New York (Brooklyn). Finally it describes the operation of the sentencing councils in these two courts and tries to assess their ability to reduce disparity.

I. THE PROBLEM

Under any system of law, unless there is but one mandatory penalty, the sentence of an offender will depend partly on the identity of the sentencing judge. Under American law, the trial judge's views and values can play a particularly important role. The law provides practically no guidelines regarding the weights to be given to the many aggravating and mitigating circumstances of the crime, the offender, and the victim. Unlike his colleagues outside the Anglo-American tradition, the American judge need not and generally does not give reasons for his sentencing decision. Since the legal sentencing frames are normally broad, especially for the more serious crimes, similar offenders who commit similar offenses under similar circumstances may receive substantially different sentences. Moreover, because American law does not allow an independent appeal of the sentence,² the trial judge's sentence is generally final, subject to adjustment only by parole or executive pardon.³

Sentence disparity is a matter of concern to the offender, the judge, and the criminal justice system as a whole. The person most directly affected is the offender. As James V. Bennett, a former director of the Federal Bureau of Prisons, observed:

The prisoner who must serve his excessively long sentence with other prisoners who receive relatively mild sentences under the same circumstances cannot be expected to accept his situation with equanimity. The more fortunate prisoners do not attribute their luck to a sense of fairness on the part of the law but to its whimsies. The existence of such disparities is among the

² The propriety of the sentence may be indirectly called into question on appeal in certain circumstances. *See generally* M. FRANKEL, *CRIMINAL SENTENCES: LAW WITHOUT ORDER* (1973).

³ *But see* ME. REV. STAT. ANN., tit. 15, §§ 2141-44 (1965), which requires state judges to give reasons for their sentences. And ten states permit some sentencing appeal. *See, e.g.*, ALASKA STAT. §§ 12.55.120, 22.10.020 (1970); ARIZ. REV. STAT. ANN. §§ 13-1717 (1956); COLO. REV. STAT. ANN. § 40-1-509 (1971). *See also* S.B. 1, 94th Cong., 1st Sess. (would allow separate appeal of a sentence).

major causes of prison riots, and it is one of the reasons why prisons so often fail to bring about an improvement in the social attitudes of their charges.⁴

Judges, too, are disturbed by these variations. At the urging of the federal judiciary, the Congress, in 1958, passed a law providing for sentencing institutes where federal judges could exchange information on sentencing alternatives and practices with the aim of reducing undesirable sentence disparity.⁵ More recently, Judge Marvin E. Frankel discussed the problem in his landmark book, *Criminal Sentences: Law Without Order*.⁶

[I]n the great majority of federal criminal cases . . . a defendant who comes up for sentencing has no way of knowing or reliably predicting whether he will walk out of the courtroom on probation, or be locked up for a term of years that may consume the rest of his life, or something in between.

The costs of disparity to the system are high. As Mr. Justice Jackson said when he was Attorney General of the United States:

It is obviously repugnant to one's sense of justice that the judgment meted out to an offender should depend in large part on a purely fortuitous circumstance; namely the personality of the particular judge before whom the case happens to come for disposition.⁷

For all concerned, sentence disparity offends a shared concept of fairness.

II. APPROACHES TO THE STUDY OF DISPARITY

In order to measure sentence disparity caused by differences among judges, all other factors that may affect the sentence must be controlled. This goal has been approached in three ways.

A. Random Assignment

If a court assigns cases at random to its judges, they will receive comparable groups of cases.⁸ Differences in sentences must then be

⁴ J. BENNETT, OF PRISONS AND JUSTICE, S. DOC. NO. 70, 88th Cong., 2d Sess. 319 (1964).

⁵ H.R. REP. NO. 1916, 85th Cong., 2d Sess. (1958).

⁶ M. FRANKEL, *supra* note 2, at 6. See also JUSTICE IN SENTENCING: PAPERS AND PROCEEDINGS OF THE SENTENCING INSTITUTE FOR THE FIRST AND SECOND UNITED STATES JUDICIAL CIRCUITS (L. Orland & H. Tyler, Jr. eds. 1974); K. DAVIS, DISCRETIONARY JUSTICE 133-41 (1969) (considers the problem in a wider context).

⁷ 1940 ATT'Y GEN. ANN. REP. 5-6.

⁸ Within the limits of the sampling error.

due to differences among the judges; such random assignment creates a natural controlled experiment. The earliest studies of sentence disparity relied on this design, but were not alert to its operational limitations.

The first of these was Everson's study in 1919, which compared the sentencing patterns of forty-two magistrates in New York City.⁹ He found that after conviction for public intoxication the frequency of suspended sentences varied from less than 1 percent to 83 percent among the magistrates; moreover, the frequency of suspended sentences tended to reflect the magistrate's ethnic background. In 1933 Gaudet et al. published their first finding from a study of 7,442 sentences imposed by six county judges of a New Jersey court that claimed to distribute its cases randomly.¹⁰ The most lenient judge ordered incarceration in 34 percent of his cases, the most severe judge for 58 percent of the offenders he sentenced. The sentencing patterns showed no change over time for three of the judges, while two became more severe and one more lenient. In 1938 Gaudet published disparity figures by type of crime.¹¹

These studies have limitations. Random assignment will produce initially comparable groups only if correctly applied across a large number of cases. In practice, considerations of court management or simple carelessness may interfere with the operation of the random assignment rule.¹² Moreover, even if assignment is random, cases are usually assigned before they reach the sentencing stage. Although sensible for the court, by the time the randomly assigned cases reach the sentencing stage, their mix is likely to differ from judge to judge. Differential rates of guilty pleas and convictions after trial, factors that are themselves partly functions of personality differences among the judges, will then affect the mix of cases and, in turn, the sentences.

B. Selecting Comparable Groups of Cases

Without prior random assignment, comparability might be

⁹ Everson, *The Human Element in Justice*, J. CRIM. L. & CRIMINOLOGY 90 (1919).

¹⁰ GAUDET, HARRIS & JOHN, *Individual Differences in the Sentencing Tendencies of Judges*, 23 J. CRIM. L. & CRIMINOLOGY 811 (1933).

¹¹ Gaudet, *Individual Differences in the Sentencing of Judges*, 32 ARCH. OF PSYCHOLOGY 1 (1938).

¹² In the federal courts, a judge will occasionally be bypassed if he is already burdened with cases requiring an unusual amount of trial time. The random assignment rule is also distorted when a new judge comes to the court and receives his first cases by transfer from his colleagues; if he came from the U.S. Attorney's Office, he will not be assigned criminal cases prepared during his time at the Office.

sought in the following manner: first, the cases of each judge are divided into subgroups according to characteristic factors likely to affect the sentence, such as the crime, the defendant, and the victim; second, comparisons are drawn among judges for cases within each subgroup. As the subgroups become more specific, they will approach comparability. The analyst can never be certain, however, that the groups are perfectly comparable because he cannot rule out the possibility that some unknown variable affecting the sentences is more heavily represented in the cases of one judge than in those of another.¹³

One of the more elaborate examples of this approach to the study of sentence disparity is Green's investigation of the Philadelphia criminal court during the late 1950s.¹⁴ Based on the 1,437 convictions obtained during a seventeen-month period in 1956-57, he analyzed the sentencing patterns of the eighteen judges in cases where the severity of the crime, the offender's criminal record, and other important factors were held constant. Green's data arrangement was appropriate, but his conclusions were not. In contrast to earlier disparity studies, he claimed to have found only minimal sentence disparity. But Green misread his data, which clearly show sentence disparity, especially in the middle ranges of severity.¹⁵

The ready availability of precise statistics for all United States federal court cases has generated several studies of sentence disparity within the federal system. One of the first comparisons of this kind was presented to the Pilot Institute of Sentencing, held in Boulder, Colorado, in the summer of 1959. One participant reported:¹⁶

The most shocking exhibit . . . was a chart showing the disparity of sentences in income tax cases over a period of 12 years. In one district all of the defendants who were convicted or pleaded guilty were sent to prison. The percentage of those sent to prison in other districts varied widely, from under 5 percent to over 90 percent.

The magnitude of the differences in the sentencing patterns of var-

¹³ Cf. D. CAMPBELL & J. STANLEY, EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH (1966); F. CHAPIN, EXPERIMENTAL DESIGNS IN SOCIOLOGICAL RESEARCH (1955); H. ZEISEL, SAY IT WITH FIGURES (5th ed. 1968).

¹⁴ E. GREEN, JUDICIAL ATTITUDES IN SENTENCING (1961) (Cambridge Studies in Criminology). See also R. HOOD, SENTENCING IN MAGISTRATES' COURTS, A STUDY IN VARIATIONS OF POLICY 123 (1962); Zeisel, *Methodological Problems in Studies of Sentencing*, 3 LAW & SOC'Y 621-33 (1969).

¹⁵ *Id.* at 134-38.

¹⁶ Thomsen, *Sentencing in Income Tax Cases*, 26 FED. PROBATION, March 1962, at 10.

ious districts is so great that it is reasonable to infer that the judges' differing sentencing philosophies are a primary cause of the disparity; this inference assumes that the mix of cases in the different districts does not vary enough to account for the observed sentence disparities. In order to test this assumption, it would be necessary to break down the mix of cases into comparable subgroups as described above. Rau has made a trial step in this direction by analyzing sentences for ten major crime categories controlling for the offender's prior record.¹⁷ He found significant disparities among the circuits. Additional disparities might have emerged had he looked at districts within each circuit.

A study now in progress by Tiffany, Avichai, and Peters, sponsored by the American Bar Foundation, promises to carry this type of analysis several steps further. They have collected detailed data on all offenders convicted in the federal system during a two year period, the crimes of which they had been convicted, and their type of plea. The authors will be able to analyze the sentencing patterns in the various districts by comparing those cases that share particular combinations of characteristics likely to affect the sentence.¹⁸ It would be useful to push the analysis still further and compile comparative data on individual judges in order to see the interplay among judges, districts, and circuits.¹⁹

While these refinements promise more specific disparity measures, the approach has natural limits. There are only a certain number of convictions in the system, and every refinement step makes each subgroup smaller and each comparison less reliable.²⁰ The sample of cases, moreover, cannot be safely expanded by including cases from more years, because sentencing patterns may change over time as judges retire and are replaced.

C. Identical Cases

A third approach to the study of sentence disparity requires different judges to "sentence" an identical group of cases, thereby ensuring absolute comparability. Because criminal sentences in

¹⁷ Rau, *Sentencing in the Federal District Courts (1972)* (mimeographed paper prepared for the National Institute of Law Enforcement and Criminal Justice (LEAA)).

¹⁸ They have published a partial analysis of their data but have not yet made comparisons by districts. See Tiffany, Avichai & Peters, *A Statistical Analysis of Sentencing in Federal Courts: Defendants Convicted After Trial, 1967-1968*, 4 J. LEGAL STUDIES 369 (1975).

¹⁹ One might extend the comparison to the state courts in the federal districts to learn whether sentencing standards in the two systems affect each other.

²⁰ Beyond a certain level of refinement, some subgroups will lose all their cases and thus lose their comparative value.

American courts are always imposed by a single trial judge, this approach must to some extent rely on simulated sentencing decisions, whereby each judge states the sentence he would give if he were the sentencing judge.²¹

The first such attempt was made in 1961 by federal judges at the Sentencing Institutes held by the Fifth Circuit in New Orleans²² and by the Sixth, Seventh and Eighth Circuits in Highland Park, Illinois.²³ The sentences handed down in the sample cases by the conference judges varied widely. In one case of income tax evasion, for example, the recommended sentences ranged from a six-month suspended prison sentence to a five-year prison sentence with a \$20,000 fine.²⁴ In an embezzlement case the sentences ranged between probation and five years custody.²⁵

The Federal Judicial Center recently conducted a systematic study of disparity using simulated sentencing.²⁶ Second Circuit judges read presentence reports and made sentence recommendations in thirty cases. Substantial disparities were found; in sixteen of the twenty cases submitted to all of the judges, there was disagreement in regard to the propriety of incarceration.

In England, Rupert Cross, throughout his career concerned with the problems of sentencing, has made informal experiments with sentencing. One of them was designed to show that judges too are human. The case involved manslaughter through grossly negligent driving. Some of the judges received the case ornamented by a legally irrelevant variable: the negligent driver was accompanied by a married woman, not his wife. Throughout, the man with the woman received a more severe sentence.²⁷

²¹ In many countries, especially on the European continent, sentences for major crimes are determined by a tribunal of several judges. Such a system would permit disparity to be measured without simulation. But this seemingly perfect "naturally controlled experiment" suffers from a potential difficulty: the unavoidable bargaining process may move some judges to make their initial recommendation artificially high or low for negotiating purposes.

²² *Sentencing Institute and Joint Council for the Fifth Circuit, New Orleans, 1961*, 30 F.R.D. 185 (1962) [hereinafter cited as *New Orleans Conference*].

²³ *Seminar and Institute on Disparity of Sentences for Sixth, Seventh, and Eighth Judicial Circuits, Highland Park, Illinois, 1961*, 30 F.R.D. 401 (1962) [hereinafter cited as *Highland Park Conference*].

²⁴ *New Orleans Conference* at 226.

²⁵ *Highland Park Conference* at 500.

²⁶ A. Partridge & W. Eldridge, *The Second Circuit Sentencing Study*, Federal Judicial Center (August 1974).

²⁷ From a personal communication to Hans Zeisel. Cross thus anticipated a favorite experimental pastime of present-day psychologists. Cf. Landy & Aronson, *The Influence of the Character of the Criminal and Victim on the Decisions of Simulated Jurors*, 5 J. EXPERIMENTAL & SOCIAL PSYCHOLOGY 141-52 (1969); Sigall & Ostrove, *Beautiful But Dangerous*:

Roger Hood, who earlier had studied sentence disparity in the English magistrate courts by the "comparable cases" method, studied sentence disparity by submitting identical cases of serious driving offenses to a panel of judges. To make the simulation more realistic, he included excerpts from the trial record, highlighted the testimony, and gave a careful description of the offender.²⁸

Despite such precautions, simulated sentencing recommendations may differ from real sentencing decisions. To impose a long prison term is a difficult task even after years on the bench; a simulated sentence sends no one to jail. Moreover, simulated sentencing is based on a written record which, however carefully prepared, cannot substitute for information gained from a sentencing judge's presence in court. It is reasonable to suppose that a less complete picture of the case leaves more to the judge's imagination and increases the likelihood of disagreements over the sentence.²⁹ In addition, the judge may suspect that the experiment is designed to measure his "severity" and, if only unconsciously, may seek to dispel an unwanted reputation by deviating from his normal sentencing practice.

The many cautionary notes we have added to the description of these earlier studies must not lead to a general distrust of them. As studies of the same problem using various research approaches proliferate, the valid ones, whatever their individual imperfections, tend to support each other by mutual corroboration.³⁰

The "identical case" approach offers the best opportunity for measuring disparity, and the sentencing council provides the optimal arrangement of this approach. While elements of simulation remain, the participating judges know that their recommendations can and often do have a real impact on the sentence actually imposed. Under present law no more realistic arrangement can be devised that will allow several judges to sentence one offender.

III. THE SENTENCING COUNCIL

The first sentencing council was started in the Eastern District of Michigan in 1960,³¹ followed by the Eastern District of New York

Effects of Offender Attractiveness and Nature of the Crime on Juridic Judgment, 31 J. PERSONALITY AND SOCIAL PSYCHOLOGY 410-14 (1975).

²⁸ R. HOOD, *SENTENCING THE MOTORING OFFENDER* (1972).

²⁹ This is probably a special case of a more general proposition, applicable to all experiments which use abbreviated materials, such as abstracted trial scripts in jury experiments.

³⁰ See generally H. ZEISEL, *supra* note 10, at 200-39 (ch. 13, Triangulation of Proof).

³¹ See Doyle, *A Sentencing Council in Operation*, 25 FED. PROBATION, Sept. 1961, at 27;

in 1962,³² the Northern District of Illinois in 1963,³³ and, more recently, the District of Oregon. All four councils are still in operation. Data from the New York³⁴ and Chicago³⁵ courts form the basis of this study.

The New York and Chicago courts generally hold weekly meetings of the sentencing council. Several days before the meeting, each participating judge receives a copy of the presentence report prepared by the court's probation office for each offender. The report contains a short description of the offense and the background of the offender. The judges then record their sentence recommendations. The cases and recommendations are discussed at the next council meeting. After hearing the discussion and the sentences recommended by his colleagues, the sentencing judge makes his final decision. The council is purely advisory; the sentencing judge retains complete discretion in making the final decision.

The Chicago and New York councils differ in important respects. In New York, every case is considered by panels of three

Hosner, *Group Procedure in Sentencing: A Decade of Practice*, 34 FED. PROBATION, Dec. 1970, at 18; Kaufman, *Sentencing: The Judge's Problem*, 24 FED. PROBATION, March 1960, at 3; Levin, *Toward a More Enlightened Sentencing Procedure*, 43 NEB. L. REV. 499 (1966); Smith, *The Sentencing Council and the Problem of Disproportionate Sentences*, 27 FED. PROBATION, June 1963, at 5; cf. Sigurdson, *The Federal Sentencing Council: An Attempt at Disparity Reduction*, 1964 (summary of unpublished master's thesis, U. of Michigan School of Social Work).

³² Zavatt, *Sentencing Procedure in the U.S. District Court for the Eastern District of New York*, 41 F.R.D. 469 (1966).

³³ B. Meeker, *Values of a Sentencing Council*, Oct. 1964 (paper presented at the Institute and Joint Council on Sentencing, Lompoc, Calif.) (The University of Chicago Law Library).

³⁴ The New York sample is composed of 624 of the cases considered by the council panels during the calendar year 1973. The judges keep the records in their personal files. Nine judges provided access to them, leaving three judges whose cases were not included in the sample; one had died, another had retired, the third was a visiting judge.

Administrative Office statistics indicate that the number of offenders sentenced during 1973 was around 1,089 (Courtesy of James McCafferty, Chief of the Operations Branch of the Administrative Office of U.S. Courts). Until 1974 sentences imposed during the summer, some 20 percent of all cases, did not come before the council. Zavatt, *supra* note 32. The 624 cases, representing twelve judges, constitute therefore approximately 57 percent of all cases. Three-fourths of the cases came before the regular three-member panels; one-fourth of the cases came before two judges. In 14 cases no record could be found of a judge's recommended sentence.

³⁵ The Chicago sample of 1,619 cases consisted of all the cases sentenced during the eighteen month period between January 1, 1972, and June 30, 1973. Of these, 518 were brought before the sentencing council; 566 belonged to judges who did not participate in the council; and 535 cases were not brought before the council by the judges who otherwise participated in its deliberations. The sample does not include 17 cases for which indictment was filed before 1971; it also omits 24 cases for which the identity of the sentencing judge could not be determined.

judges: the sentencing judge and two of his colleagues assigned more or less in rotation. In Chicago council participation is voluntary, and each case comes before the full council of participating judges.³⁶

This study measures the degree of existing disparity in both courts, discusses the ability of each council to reduce disparity, and evaluates the effects of the organizational structures of the two councils. First, however, it will be necessary to define sentence disparity more precisely.

IV. THE MEASUREMENT OF DISPARITY

Sentence disparity can be measured at various levels. The first level of disparity is disagreement among judges over the type of sentence—custody (prison) or noncustody (probation or fine). At a second level, differences in the duration of recommended sentence can be measured within each type. Finally, disparity may be measured on a scale that considers the duration as well as the type of sentence. Table 1 gives an overview of the sentences at the disposal of a federal judge:³⁷

Table 1

SIX MAJOR SENTENCING FRAMES FOR FEDERAL CRIMES

	<u>Sentencing Frames</u>		<u>Examples of Crimes in that Frame</u>
	<u>Minimum</u>	<u>Maximum</u>	
1	Probation	\$1,000 fine and/or 1 year	a) Mail theft of property up to \$100 (18 U.S.C. § 1708) b) Embezzlement by employee of bank—up to \$100 (18 U.S.C. § 656)
2	Probation	a) 3 years b) \$1,000 fine and 3 years	a) Destruction of letter boxes or mail (18 U.S.C. § 1705) b) Impersonating a federal officer (18 U.S.C. § 912)
3	Probation	a) \$2,000 fine and 5 years b) \$5,000 fine and 5 years c) \$10,000 fine and 5 years	a) Mail theft of property over \$100 (18 U.S.C. § 1708) b) Sale or receipt of stolen vehicles (18 U.S.C. § 2313) c) Selective Service violation (50 U.S.C. § 462(a))

³⁶ Originally, the sentencing council in Chicago consisted of two panels of six members each. Eventually, one panel stopped meeting, and the members who wished to remain on the council joined the other panel. At the time of this study, nine members of the court were participating in council activities, though not all of them all of the time.

³⁷ A few rare offenses carry different upper limits, *e.g.*, 18 U.S.C. § 701 (1970) (illegal use of a government insignia); 18 U.S.C. § 3 (1970) (violation of a government park regulation); 18 U.S.C. § 114 (1970) (\$1,000 and/or seven years for maiming).

4	Probation	a) \$5,000 fine and/or 10 years	a) Theft from interstate shipment over \$100 (18 U.S.C. § 659)
		b) \$10,000 fine and 10 years	b) Interstate transport of stolen goods (18 U.S.C. § 2314)
5	Probation	a) \$5,000 fine and 20 years	a) Bank robbery with force (18 U.S.C. § 2113(a))
		b) \$10,000 fine and 20 years	b) Extortion (18 U.S.C. § 894)
6	10 years	a) Life imprisonment	a) Murder or kidnapping during commission of bank robbery (18 U.S.C. § 2113(e))
		b) \$100,000 fine and life imprisonment	b) Narcotics violation indicating a continuing criminal enterprise (21 U.S.C. § 848(a))

A. Levels of Disparity

1. *Type of Sentence.* A federal judge can almost always choose between ordering imprisonment or probation. He also can mix the two by imposing a short prison term of six months or less, followed by a longer period of probation.³⁸ In most cases he can also impose a fine, and sentences for more than one count may be imposed concurrently or consecutively.

The most crucial part of the sentencing process is the decision whether the offender is to be imprisoned. Stability of job and family life are at stake, even if the sentence is only a short one. This first level of disparity, between custody and noncustody, is reflected in Table 2.

Table 2

AGREEMENT AND DISAGREEMENT ON THE ISSUE OF CUSTODY
(AMONG THREE JUDGES)

	<u>Chicago*</u> %	<u>New York</u> %
Agree to impose no custody	27.8	11.4
Agree to impose custody	40.8	58.9
Total agreement	68.6	70.3
Disagree on whether to impose custody	31.4	29.7
	100%	100%
	(439)	(460)

*simulated for three-judge panels

³⁸ See 18 U.S.C. § 3651 (1970).

The two courts show almost identical figures on this first level of disparity.³⁹ In 30 percent and 31 percent of their respective cases, the judges of the New York and Chicago councils disagree whether the offender should be incarcerated. The distribution of the cases in which they agree, however, reveals that the sentencing pattern is more severe in the New York court. The proportion of agreed-upon noncustody cases is 11 percent in New York; it is 28 percent in Chicago. The custody cases show the reverse pattern.

2. *Duration of Sentence.* The second level of potential disparity is disagreement over the length of the sentence, as shown in Table 3.

Table 3

AGREEMENT AND DISAGREEMENT ON DURATION OF SENTENCE

	Custody Sentences		Noncustody Sentences	
	Chicago*	New York	Chicago*	New York
Agree on Duration	12%	10%	98%	82%
Disagree on Duration	88%	90%	2%	18%
	100%	100%	100%	100%
Percent of all Council Cases in which there was agreement on type of sentence (See Table 2)	(40.8%)	(58.9%)	(27.8%)	(11.4%)

*Simulated for three-judge panels

If all judges agree on custody, there will be disagreement over its length in nine out of ten cases, both in New York and in Chicago. If all favor probation, there will be disagreement over its length in New York 18 percent of the time, in Chicago only 2 percent of the time.

3. *Disparity as to Type and Duration of Sentence.* Measuring disparity for both type and duration of sentence requires a yardstick that converts both dimensions to a common measure. The Administrative Office of the United States Courts has developed a scale for that purpose which we have slightly modified to provide finer detail at the upper end of the scale and to reduce the importance of differences in length of probation (Table 4).

³⁹ Because the incidence of agreement is artificially affected by the number of participating judges, the Chicago disparity figures were computed by simulating the three-judge panels of the New York court. This was done by randomly selecting two judges from the group of consulting judges that participated in the particular case, and treating them as the two advisors to the sentencing judge.

Table 4

POINT SCALE OF SENTENCE SEVERITY

	<u>Developed by the Administrative Office of the U.S. Courts</u>	<u>As Modified</u>
Fine	1	1
Probation (months)		
1-12	1	1
13-36	2	2
Over 36	4	<u>3*</u>
Split-Sentence (Jail + Probation)	4	4
Prison (months)		
1-6	3	3
7-12	5	5
13-24	8	<u>7</u>
25-36	10	<u>9</u>
37-48	12	<u>11**</u>
49-60	14	<u>13</u>
61-120	25	<u>15</u>
over 120	50	<u>17</u>

(Add two points for every year.)

*underlined figures are our modifications

**following the example of the Federal Judicial Center study (note 23 *supra*), custody sentences under the Youth Corrections Act count as four year terms.

Even as modified, the scale fails to reveal minor differences among sentences. For example, both two and three years of probation have a point value of 2; a fine and one year of probation both have a value of 1. Moreover, the scale provides no information on whether special conditions, such as participation in a drug-abuse program, are attached to the sentence. On the whole, therefore, the scale will understate slightly the differences among sentences.

We propose to define and measure disparity as follows:

Sentence disparity is the expected percent difference between two sentences if two judges, randomly selected from the court, were to sentence the same case independently. Their two sentences are expressed as a percentage of their common mean.

Using the modified scale, we can derive the disparity figures from the sentencing council data. The measure can, of course, be applied to the disparity among more than two judges, by averaging the disparities between all possible combinations of two judges. The following example is for the case of three judges, and thereby describes the computation method for the New York council (Table 5).

Table 5
EXAMPLE FOR COMPUTING
SENTENCE DISPARITY AMONG THREE JUDGES

	Example		
	Judge	Sentence	Points
(1) Translate the sentences recommended by each judge into their respective point values:	A	3 yrs.	9 pts.
	B	5 yrs.	11 pts.
	C	7 yrs.	13 pts.
(2) Compute mean sentence		(33 ÷ 3 =)	11pts.
	Judge Combinations		
(3) Compute point differences between each pair of judges:	A & B		2 pts.
	A & C		4 pts.
	B & C		2 pts.
(4) Compute mean difference		(8 ÷ 3 =)	2.67 pts.
(5) Express mean difference as percentage of mean sentence		(2.67 ÷ 11 =)	24 percent

The sentence disparity in the example given in Table 5 is 24 percent. Averaging the disparity measures in a great number of cases from a court in which all judges participate in council panels yields a disparity measure for the court.

It would be possible to express disparity in terms of the absolute point difference between two judges, by omitting operation (5). The percent measure adopted here, on the other hand, considers the sentence difference between one and three years to be more important than the difference between ten and twelve years,⁴⁰ even though the absolute disparity is two years in either case. The percent measure also corresponds to an important practical difference between these two cases: for the parole board, which ultimately determines how much time the offender will actually serve, the difference between one and three years is undoubtedly more critical than that between ten and twelve years.

The disparity measures for the two courts, as derived from the council data, are presented in Table 6 (page 123). Two judges will thus differ, on the average, by between one-third and one-half of the mean sentence.

⁴⁰ The scale of the Administrative Office already makes this adjustment to a limited extent; in the lower prison sentence ranges one point represents four months, in the higher ranges, six months.

Table 6

AVERAGE SENTENCE DISPARITIES OF THE TWO COURTS

Chicago*	36.7 percent (Std. Dev. 33.0%)
New York	45.5 percent (Std. Dev. 22.9%)

*for the one-third of the cases brought before the council

B. Differences in Severity Among Judges

One cannot necessarily infer from the existence of sentence disparity that some judges are generally more severe than others. Even if sentence disparity were present, it might be true that all judges sentenced, on the average, with equal severity. Each judge might simply fluctuate around the same average. This situation does not exist, however, in either of the two courts studied; in both, some judges are clearly more severe than others.

The differences in severity among judges can be shown by measuring the deviation of each judge's sentencing recommendation from the mean of all sentences proposed in any one case, taking into account whether the deviation was toward the severe (+) or the lenient (-) side. Since each judge takes part in numerous council deliberations in all possible combinations with his colleagues, a reliable average deviation for each judge, in all cases in which he participates, may be computed:

Table 7

SEVERITY OF JUDGES*

Chicago			New York		
Judge	Percent Deviation	Number of Cases	Judge	Percent Deviation	Number of Cases
A	-11	(329)	A	- 5	(172)
B	+10	(311)	B	-12	(201)
C	+ 7	(293)	C	-11	(104)
D	+ 4	(312)	D	+ 7	(164)
E	- 1	(385)	E	-11	(110)
F	-10	(331)	F	+21	(128)
G	+ 5	(161)	G	+24	(59)
H	+ 3	(48)	H	-10	(28)
			I	-21	(61)
			J	+58	(27)
			K	+20	(77)
			L	- 2	(5)

*percent average deviation from the mean recommendation of all judges

If a judge always stood at his "average severity" position (which, of course, he does not), Chicago Judge *B* would always be 10 percent more severe than the average, and Judge *F* would always be 10 percent less severe than the average. Their sentences would vary between 110 percent for Judge *B* (100 + 10) and 90 percent for

Judge *F* (100–10), a difference of about 20 percent. But none of the judges on either court sentence with consistent severity, as, for example, the New York voting pattern illustrates. Hence, in any given case the difference between any two judges will, as a rule, be larger or smaller than their average severity figures would indicate (Table 8).

Table 8
VOTING POSITION OF NEW YORK JUDGES*

	Judge						
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>	<u>K</u>
	%	%	%	%	%	%	%
least severe	22	26	22	14	28	10	17
shares least severe position with at least one other judge	24	21	24	15	19	8	8
all judges agree or the sentencing judge is in the middle	24	28	33	26	33	24	21
shares most severe position with at least one other judge	13	14	7	17	12	18	8
most severe	<u>17</u>	<u>11</u>	<u>14</u>	<u>28</u>	<u>8</u>	<u>40</u>	<u>46</u>
	100	100	100	100	100	100	100

*with 75 or more cases

Therefore, it is not simply the fluctuation of sentences that causes disparity; some judges are in fact more severe than others.⁴¹ Such differences in sentencing philosophies appear to be a major cause of the sentence disparity that sentencing councils are designed to reduce.

V. COUNCIL EFFORTS TO REDUCE DISPARITY

The primary purpose of the sentencing council is to reduce sentence disparity by confronting the sentencing judge with the views of his colleagues. If disparity is revealed, it is assumed that the sentencing judge will change his sentence to reduce the disparity. To test this assumption, it is necessary to consider, first, how often and under what circumstances the sentencing judge alters his

⁴¹ Actually, the situation is somewhat more complex. Judges will vary in their relative severity level according to the particular type of case. These patterns will be examined in a separate article along with the case characteristics that appear to promote disparity and disparity reduction.

original sentence decision and, second, the extent to which such changes reduce disparity.

As Table 9 shows, sentencing judges in both courts change their sentences in only about one-third of the cases they bring before their councils. Roughly half of these changes involve shifts in sentence type; the other half alter sentence length.

Table 9
CHANGES BY THE SENTENCING JUDGE
BETWEEN RECOMMENDATION AND DISPOSITION

	<u>Chicago</u> %	<u>New York</u> %
Increases Sentence	13	7
in type	(4)	(3)
in duration	(9)	(4)
Reduces Sentence	20	35
in type	(11)	(20)
in duration	(9)	(15)
	<hr/>	<hr/>
Total Changes	33	42
No Change	67	58
	<hr/>	<hr/>
	100%	100%
	(N = 434)*	(N = 460)*

*Cases with at least 3 judges in which the sentencing judge recorded an initial recommendation

Table 9 also shows that in both courts sentence reductions outnumber sentence increases. New York shows a higher frequency of change (42 percent) than Chicago (33 percent). One might have expected the participating judges in Chicago to be more amenable to change, since participation in this council is voluntary. On the other hand, Chicago judges might change less than the New York judges because initial disparity was lower in Chicago. Another interesting difference is that the imbalance between increases and reductions is more pronounced in New York (7 to 35) than in Chicago (13 to 20). It will be shown, however, that the major part of this greater imbalance in New York is unrelated to the council.⁴²

Under what circumstances, then, will the judge change his original position? It is reasonable to assume that the extent of his disagreement with his fellow judges will be a critical factor in determining whether the sentencing judge changes his sentence. Table 10

⁴² See text at note 44 *infra*.

Table 10
COUNCIL VOTE AND SUBSEQUENT CHANGES

Action of Sentencing Judge	COUNCIL VOTE AND SUBSEQUENT CHANGES						Total
	(1) Position of Counseling Judges All judges are more severe (++) %	(2) Some judges are more severe, at least one shares the trial judge's view (+) %	(3) All judges agree with sentencing judge (=) %	(4) Judges disagree in both directions (+-) %	(5) Some judges are less severe, but at least one shares the trial judge's view (-) %	(6) All judges are less severe (--) %	
<u>Chicago</u>							
becomes more severe	46	17	9*	10	2*	2*	13
does not change	50	77	88	67	62	24	67
becomes less severe	4*	6*	3*	23	36	74	20
Share of all Cases	100 (25) (6%)	100 (149) (34%)	100 (33) (8%)	100 (129) (29%)	100 (60) (14%)	100 (38) (9%)	100 (N=434) %
<u>New York</u>							
becomes more severe	28	7	.	8	.	.	7
does not change	57	68	77	62	54	28	58
becomes less severe	15*	25*	23*	30	46	72	35
Share of all Cases	100 (16%)	100 (20%)	100 (13%)	100 (20%)	100 (14%)	100 (17%)	100 (100%) (N=460)

*unexplained changes

supports this assumption. The six columns for each court represent the six possible positions in which the sentencing judge may find himself in relation to his colleagues.⁴³

In the Chicago court (Table 10), if all the counseling judges favor a more severe sentence (col. 1), the sentencing judge will increase his sentence in 46 percent of the cases. If some judges recommend a higher sentence, but at least one shares the sentencing judge's recommendation (col. 2), the judge will increase his sentence in only 17 percent of the cases. On the other hand, if all judges vote for a more lenient sentence (col. 6), the sentencing judge will reduce the severity of his sentence in 74 percent of the cases; he will reduce the sentence 36 percent of the time even if only some of his colleagues vote for a more lenient sentence as long as none votes for a higher one. Finally, the sentencing judge may move in either direction if the counseling judges split and recommend sentences both higher and lower than his original recommendation. The pattern of the New York court is similar, even though that court has a more dramatic imbalance between sentence reductions and sentence increases.

Table 10 shows the frequency and direction of changes by the sentencing judge after the case has been before the council. Two types of changes cannot plausibly be attributed to the council. The first type (marked with an asterisk) are changes where the sentencing judge moves in the opposite direction from that recommended by the counseling judges. If none of the counseling judges favor a more lenient sentence (cols. 1, 2, 3) but the sentencing judge reduces the sentence anyway, or if none recommend a more severe sentence (cols. 3, 5, 6) but the sentencing judge nevertheless increases the sentence, the council cannot have caused the change.

Second, there are cases in which the sentencing judge, although changing in the direction recommended by his colleagues, moves beyond the position held by the most extreme counseling judge. In one case, for instance, the sentencing judge proposed a three-year prison term, his colleagues recommended two years and one year respectively, and the judge ultimately imposed a six-month sentence. Table 11 treats both of these kinds of changes as if the sentenc-

⁴³ By subdividing the sample into groups based on the initial location of the sentencing judge, the possibility of regression—that is, of the greater likelihood of extreme values to move toward the mean upon a second measurement—arises. The role of regression in this study is difficult to determine, however, since the very purpose of the council is to pull extreme values back toward the mean. The second measure is thus not only independent of the first, but the council is most likely to reduce disparity precisely where the regression phenomenon is most likely to occur.

Table 11
MAXIMUM FREQUENCY OF COUNCIL-INDUCED CHANGES

Action of Sentencing Judge	Position of Counseling Judges		(1)	(2)	(3)	(4)	(5)	(6)	Total
	All judges are more severe (++)	All judges agree with sentencing judge (=)	Some judges are more severe, at least one shares the trial judge's view (+)	All judges agree with sentencing judge (=)	Judges disagree in both directions (+-)	Some judges are less severe, but at least one shares the trial judge's view (-)	All judges are less severe (---)		
	%	%	%	%	%	%	%	%	%
<u>Chicago</u>									
becomes more severe	42	-	17	-	10	-	-	-	11
does not change	58	100	83	100	70	72	29	73	
becomes less severe	-	-	-	-	20	28	71	16	
	100	100	100	100	100	100	100	100	
<u>New York</u>									
becomes more severe	28	-	5	-	8	-	-	-	7
does not change	72	100	95	100	69	66	36	73	
becomes less severe	-	-	-	-	23	34	64	20	
	100	100	100	100	100	100	100	100	

ing judge made no change in his sentence and thereby represents as changes only those shifts that could have been induced by the council.⁴⁴ The figures it presents, moreover, represent the estimates of the maximum frequency of council-related changes.

In both cities, the maximum frequency of council-related change is 27 percent. In 73 percent of the cases, the council did not cause the judge to change. Sentence reductions in Chicago (16 percent) and increases (11 percent) are fairly balanced; in New York, reductions occur three times as often as increases.

We have thus far assumed that nearly all decreases in columns 4, 5 and 6 are due to the council's influence. But since unexplained reductions occurred in columns 1, 2 and 3, it is likely that at least some of the reductions in columns 4, 5 and 6 are also noncouncil-related. We estimate the frequency of these unrelated changes by computing the mean percent of unexplained reductions from columns 1, 2 and 3 and applying the resulting figures (5 percent in Chicago and 21 percent in New York) to the remaining columns. The procedure is reversed to obtain the unrelated increases, which only occur in Chicago:

Table 12
"UNEXPLAINED" SENTENCE CHANGES

	Chicago	New York
Reductions*	5%	21%
Increases**	4%	0

*mean of reductions in columns 1, 2, and 3 of Table 10

**mean of increases in columns 3, 5, and 6 of Table 10

Removing these additional unrelated changes yields an estimate of the minimum frequency of council-induced changes, represented in Table 13.

Somewhere between the maximum figures in Table 11 and the minimum figures in Table 13 lies the true frequency of council-induced changes. In Chicago the sentencing judge changes as a result of the council in between 25 and 27 percent of the cases. In New York the range is between 20 and 27 percent.

⁴⁴ Conceivably, there could be an indirect relation to the council; the sentencing judge may at times place an artificially severe sentencing proposal before the council intending to make a later reduction in order to deter protracted disagreement. There is some indirect statistical evidence to that effect: "noncouncil-related" reductions are slightly more apt to occur if the counseling judges are on the severe side.

Table 13
MINIMUM FREQUENCY OF COUNCIL-INDUCED CHANGES

Position of Counseling Judges	Chicago		New York		
	(4)	(5)	(4)	(5)	(6)
Action of Sentencing Judge	Judges disagree in both direc- tions (+--)	Some judges are less severe, but at least one shares the trial judge's view (-)	Judges disagree in both direc- tions (+-)	Some judges are less severe, but at least one shares the trial judge's view (-)	All judges are less severe (--)
	%	%	%	%	%
becomes more severe	6	-	8	-	7
does not change	76	69	83	75	49
becomes less severe	18	31	9	25	51
	100	100	100	100	100
	Total	Total	Total	Total	Total
	%	%	%	%	%

*The Chicago "more severe" figures are altered because of the unexplained increases in columns 3, 5, and 6 (Table 10).

Given this estimate of the frequency of change: by how much does the sentencing council reduce sentence disparity?

VI. MEASURING DISPARITY REDUCTION

Sentence disparity has been operationally defined as the expected percent difference between two sentences if two judges, selected at random from the court, were to sentence the same case independently. To measure the extent to which sentencing councils reduce disparity under this definition would require data about how the council changes the views of the counseling judges as well as that of the sentencing judge. In the absence of reliable data on this point⁴⁵ we propose to measure disparity reduction indirectly.

Considering the following model of a sentencing council procedure:

Every sentencing case is brought before the council, which consists of all judges of the court. After the sentencing judge has effectively shared all of the relevant information on the case with his brethren, each judge independently recommends the sentence he considers appropriate. Under these conditions of complete and equal information and responsibility, the court accords equal weight to each recommendation. *The judges have agreed that the mean of the recommended sentences will become the sentence of the court.*

Such a system would not necessarily remove disparity between individual judges. But in the important sense of whether two offenders under identical circumstances would receive the same sentence, this procedure would remove disparity. To the offender it does not matter whether his sentence came about because every judge thought it to be the best sentence or whether they simply agreed on the procedure of averaging their sentences. This procedure would effectively remove disparity between any two judges, since each judge, irrespective of his own position, would have to impose the same sentence—the mean of all recommended sentences.

While the adoption of such a council rule would not be without

⁴⁵ In Chicago, the counseling judges are asked to write down their final recommendations at the end of the meeting. Those judges change their recommended sentences about half as frequently as the sentencing judge does. We were privately informed, however, that the second notations by the counseling judges are not always accurate gauges of their true opinions. Perhaps they hope to minimize intra-court disharmony by reducing the apparent disparity.

The New York court, unlike its counterpart in Chicago, does not sit en banc but rather in three-judge panels. The mean sentences of these panels will deviate somewhat from the mean sentence the entire court would have imposed. That difference is a disparity invisible to the sentencing judge because he can only know the panel's mean. Even if he accedes to that mean, the New York judge will not have removed all disparity relative to the court's mean. Statistical theory allows us to compute the size of that invisible difference (2b in Table 15),⁴⁷ and, by adding it to the visible one, it is possible to estimate the deviation of the New York sentencing judge from the mean sentence of the entire court. Table 15 compares the deviations from the mean in the two courts.

The basic disparity is higher in New York (45.5 percent) than in Chicago (36.7 percent). The lower figure for the Chicago council probably stems from its voluntary character; the judges who choose to participate in the council are relatively likeminded. The visible deviation of the sentencing judge from the mean of the council is the same in both courts (23.3 and 23.5 percent), but once the "invisible" deviation is added to the visible deviation from the mean in the New York council, the latter shows a greater total deviation.

The judges in the two courts studied, unlike those in the hypothetical sentencing council, seldom impose a sentence that is the mean of all council judges' recommendations. At times they remove only part of their deviation from the mean, and at times they even increase the deviation by moving away from the mean. The following diagrams illustrate the four types of moves open to the sentencing judge:

⁴⁷ There are 56 possible ways of selecting 3 judges from a court of 8 judges. The mean deviation for each of the 56 three-judge panels could be calculated theoretically, but in reality the only calculation that can be made is for the combination that actually occurs in a particular case. The average of the 56 mean deviations is different from the mean deviation of all 8 of the judges in the council. How is the mean deviation of a three-judge subgroup related to the mean deviation of all 8 judges? Averaged over sentences that are normally distributed, the average of the 56 mean deviations is smaller than the mean deviation of all 8 judges and accordingly an upward adjustment of 15 percent must be made in the mean deviation for the three-judge panels. For a panel of n judges and subgroups of size k , the upward adjustment is:

$$\left[\left(\frac{n-1}{n} \right)^{1/2} / \left(\frac{k-1}{k} \right)^{1/2} - 1 \right] \times 100\%$$

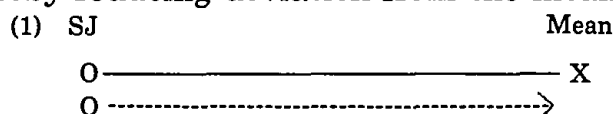
Table 15

RELATIONSHIP BETWEEN BASIC DISPARITY AND DEVIATION OF THE SENTENCING
JUDGE FROM THE MEAN IN THE TWO COURTS

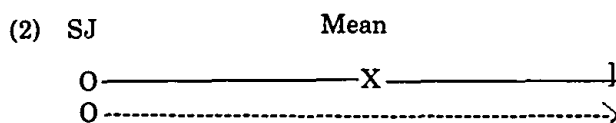
	<u>Chicago</u>	<u>New York</u>
<u>Basic Disparity (1)</u>		
Average difference between the sentence imposed in the same case by any two judges randomly selected from the court	36.7%	45.5%
<u>Visible Part of the Deviation (2a)</u>		
Distance of sentencing judge from the mean of all participating council judges	23.5%	23.3%
<u>Invisible Part of the Deviation (2b)</u>		
If the council does not consist of the entire court but only of smaller subgroups, the difference between the means of those subgroups and the (potential) mean of the entire court in each case is invisible. For 3-judge panels that difference is on the average 15% of the visible deviation	-*	3.5%
<u>Total Deviation (2a + 2b)</u>		
Total distance of sentencing judge from the mean of all judges of the court	23.5%	26.8%

*In the Chicago council all participating judges sit *en banc*.

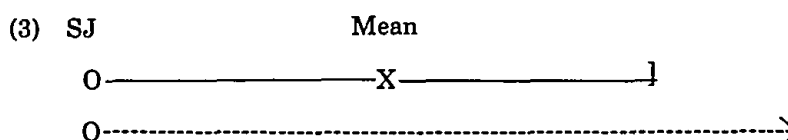
(a) The judge moves toward the mean and stops either at or before it, thereby reducing deviation from the mean.



(b) The judge moves beyond the mean but stops at a point closer to the mean than was his original recommendation, thereby achieving a net reduction of deviation.



(c) The judge moves beyond the mean and stops at a point at least as far from the mean as was his original recommendation. He thus increases deviation or leaves it unchanged.



(d) When the sentencing judge occupies the middle position, he is often so close to the mean that in changing his sentence he moves away from it, thereby increasing disparity.

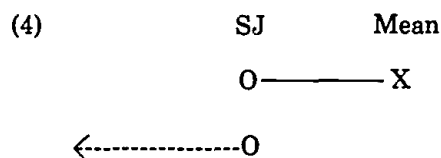


Table 16 shows the frequency of the four types of moves by the sentencing judges on each court.

The majority of these moves reduce sentence disparity, yielding a net reduction of deviation (Table 17). The sentence changes shown in Table 17 are those developed in Table 11 and hence represent the maximum estimate of changes attributable to the council.

For the 73 percent of the cases in which the sentencing judge does not change his original recommendation, the average deviation in each court is 20 percent. In the 27 percent of the cases in which the judge does change, his original deviation averages 33 percent in Chicago and 36 percent in New York, from which the changes remove 9 percentage points (of 33 and 36 percent, respectively), or about one-quarter of the deviation (27 and 25 percent, respectively).

Table 16
 FREQUENCY OF THE FOUR TYPES OF MOVES BY
 THE SENTENCING JUDGE IN THE COUNCIL

	<u>Chicago</u> %	<u>New York</u> %
<i>Reduces deviation by moving</i>		
(1) toward the mean	24	24
(2) beyond the mean, still reducing deviation	<u>33</u>	<u>32</u>
	57	56
<i>Increases deviation by moving</i>		
(3) so far beyond the mean, that deviation is increased or unchanged	31	16
(4) away from the mean	<u>12</u>	<u>27</u>
	43	44
	<u>100%</u>	<u>100%</u>

Table 17
 REDUCTION OF THE SENTENCING JUDGE'S VISIBLE DEVIATION *
 (IN PERCENT OF THE MEAN SENTENCE)

<u>Chicago**</u>	Share of Cases %	(1) Average Size of Deviation	(2) Reduction in Council	(2) as Percent of (1)
<i>Sentencing Judge</i>				
Does not change	(73)	20	0	0%
Changes	<u>(27)</u>	<u>33</u>	<u>9</u>	<u>27%</u>
Total	(100%)	23.5	2.5	<u>10.6%</u>
 <u>New York</u>				
<i>Sentencing Judge</i>				
Does not change	(73)	20	0	0%
Changes	<u>(27)</u>	<u>36</u>	<u>9</u>	<u>25%</u>
Total	(100%)	23.3	2.5	<u>10.7%</u>

*From the mean

**Only for the one-third of the cases coming before the Chicago council

The initial average visible deviation in each city was about 23 percent. In each city, these figures were reduced as a result of the sentencing council by 2.5 points, or a little over 10 percent. The drops from 27 and 25 percent to roughly 10 percent in the last column reflect the fact that the council fails to move the judge in 73 percent of the cases.

We now apply the share of the deviations removed by the council to the basic disparity⁴⁸ and thereby arrive at the final evaluation of the council's effectiveness in reducing disparity (Table 18).

Table 18

EFFECT OF THE COUNCIL ON SENTENCE DISPARITY
(AVERAGE DIFFERENCE BETWEEN THE SENTENCES
IMPOSED IN THE SAME CASE BY ANY TWO JUDGES,
RANDOMLY SELECTED FROM THE COURT)

	<u>Chicago</u>	<u>New York</u>
Disparity prior to council deliberation	36.7	45.5
Percent removed by council	-10.6	-9.6*
Disparity remaining after council	32.8	41.1

*adjusted by reducing the 10.7 share (Table 17) by 10 percent for "invisible deviation" in New York (Table 15)

In each court the council removed about 10 percent of the existing disparity, thus reducing disparity in New York from 45 to 41 percent, in Chicago from 37 to 33 percent.

VII. COMPARING THE TWO COURTS

The amount of disparity reduction is strikingly similar in the two courts. But this similarity should not mask the different processes that produced the result. In order to assess the relative impact of the two sentencing councils, their structure and operation must be compared in terms of case composition, council size, and whether participation was voluntary.

A. Differences in Case Composition

The urban character and the population of the New York and Chicago districts are roughly comparable but the composition of the cases that come before the two courts differ in a number of ways, as Table 19 shows.

⁴⁸ The ratio of these deviations to the disparity in those courts, .59 in New York and .64 in Chicago, corresponds well with the theoretical value. *Cf.* note 47 *supra*.

Table 19
DISTRIBUTION OF OFFENSES

	<u>Chicago</u>	<u>New York</u>
	%	%
Narcotics	7.8	29.2
Marijuana and Other Drugs	4.3	9.0
Bank Robbery	3.0	8.3
Theft and Interstate Transport of Stolen Goods	28.1	12.9
Other Property Crimes	19.5	11.2
All Other Crimes	37.3	29.4
	<u>100.0%</u>	<u>100.0%</u>

Over 38 percent of all New York cases were drug-related, compared to 12 percent in Chicago. Chicago also had fewer bank robberies than New York, perhaps because Illinois does not permit branch banking and hence reduces the opportunities for committing that crime. The Chicago court, on the other hand, had a higher frequency of theft, fraud, and interstate transport of stolen property, probably because Chicago is centrally located and has numerous road and railway transport systems passing through it.

The crime mix for which the New York offenders were convicted was somewhat more serious. Taking the maximum sentence the law allows for a given offense as a measure of seriousness, the average of these maximum years was 11.2 years for New York and 10.0 years for Chicago.⁴⁹ This difference in seriousness of the crimes may account for some of the differences observed between the two courts.

B. Council Size

The New York council meets in three-judge panels on a rotation basis, whereas the Chicago council meets en banc. Because the Chicago council is much larger than the New York panel, two countervailing factors may influence the sentencing judge. First, the sentencing judge may be more likely to heed the council's advice if he is opposed by a larger number of his colleagues. In New York, a

⁴⁹ The recommendations and sentences of the sentencing judges reflect this difference. The average recommendation was 8.5 in New York and 5.1 in Chicago; the average sentence in New York was 7.4 and in Chicago it was 4.9. The size of the difference suggests that factors other than the severity of the case may contribute to the different severity levels in the two courts. New York also has a higher share of convictions after jury trial (16%) than Chicago (10%). Bench trial convictions are more frequent in Chicago (5%) than in New York (2%).

maximum of two judges can oppose the sentencing judge; in Chicago, many more may oppose him (Table 20).

Table 20

LIKELIHOOD OF SHIFT DEPENDING
ON THE VOTING POSITION OF THE SENTENCING JUDGE

Position of Sentencing Judge	<u>Chicago</u>	<u>New York</u>
All judges more severe	42%	28%
Some judges more severe, none less severe	13%	7%
Judges on both sides	24%	17%
Some judges less severe, none more severe	31%	25%
All judges less severe	69%	51%

In each voting position, the Chicago judge is more likely to change than his New York counterpart. On the other hand, as the size of the council becomes larger, the sentencing judge will find himself less frequently in an extreme position, as shown in Table 21:

Table 21

DISAGREEMENT CONSTELLATION VIS-À-VIS THE SENTENCING JUDGE

	<u>Chicago</u>	<u>New York</u>
	%	%
all more severe	5	16
some more severe	34	20
all agree	8	13
disagree in both directions	30	20
some less severe	14	14
all less severe	9	17
	<u>100%</u>	<u>100%</u>

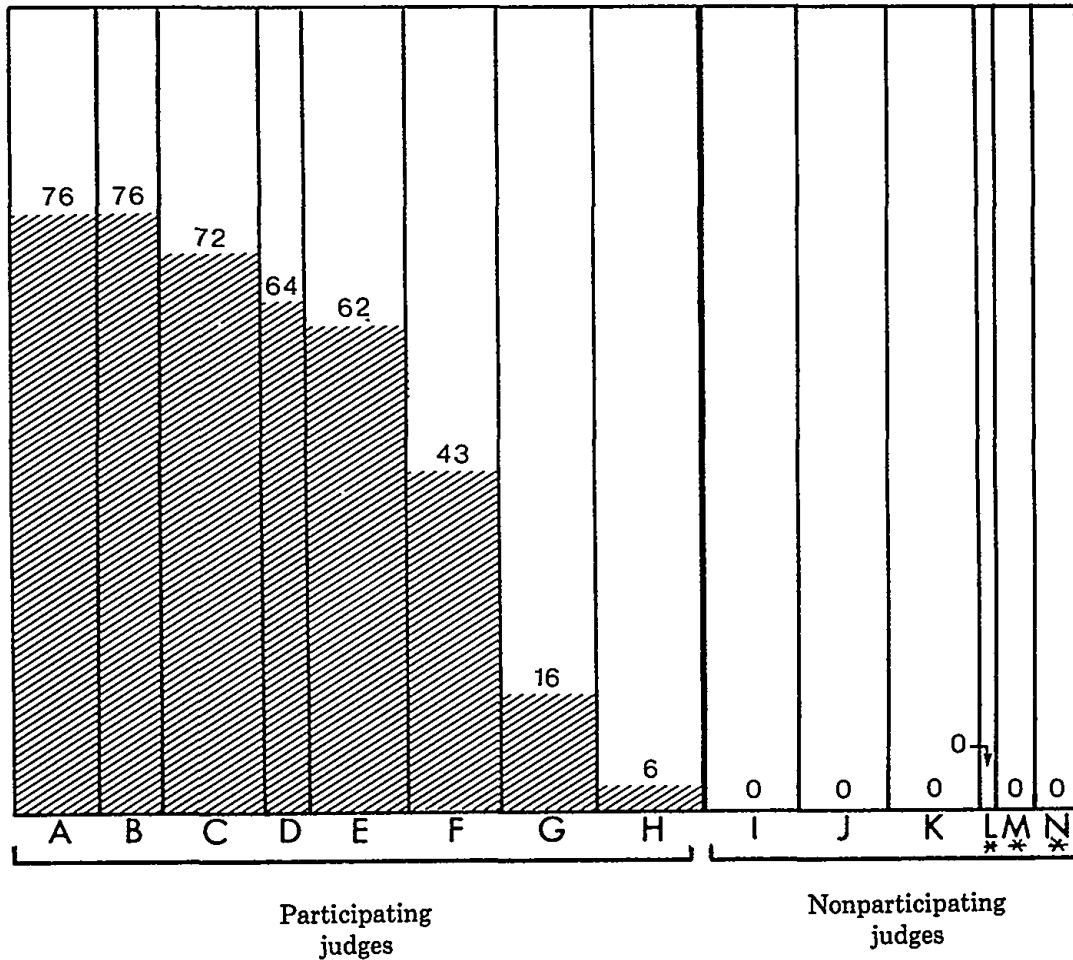
The Chicago judge is less frequently in an extreme position, but when he finds himself there he may feel more pressure to alter his original recommendation. It may well be that the countervailing effects produced by the larger council size of the Chicago court simply cancel each other.

C. Voluntary Participation

Participation in the Chicago council is voluntary, and some judges do not take part. Even those who do participate do not bring all their cases before the council. Table 22 gives an overview of this

limited participation, and Table 23 breaks down the total number of cases sentenced by the six judges who do not participate in the council and by the eight who do.

Table 22
 PERCENT OF CASES BROUGHT BEFORE THE CHICAGO COUNCIL
 BY EACH OF THE 14 JUDGES OF THE COURT



The width of the bar indicates the judge's share of cases in the sample.

*senior judges with reduced caseloads

There are thus three groups of cases that reach the sentencing stage, each accounting for about one-third of the total.

Table 23

SHARE OF THE CHICAGO CASES BROUGHT BEFORE THE COUNCIL

	Percent
Cases not brought before council	68
cases of participating judges	(33)
cases of nonparticipating judges	(35)
Cases brought before council	32
	<hr/> 100%

number of cases (1619)

The voluntary and limited participation in the Chicago council raises certain questions. First, what distinguishes the nonparticipating from the participating judges? Second, what distinguishes the cases withheld by the participating judges from those brought before the council? Finally, how much further would disparity in Chicago be reduced if council participation were obligatory? Table 24 helps answer the first question: one characteristic that distinguishes nonparticipating judges from the participating judges is their age.

Table 24

AGE AND YEARS ON THE BENCH
OF PARTICIPATING AND NONPARTICIPATING JUDGES

Participating Judges		Nonparticipating Judges	
Age	Years on Bench	Age	Years on Bench
46	*	66	5
49	*	66	8
51	1	68	9
54	*	69**	31
55	1	71	10
58	10	75**	20
60	10	76**	19
63	6		
		<hr/> Average 70	<hr/> 15 yrs.
<hr/> Average 54	<hr/> 3 yrs.		

*Less than 1 year

**Senior judges

There is a clear division: all judges under 65 participate; judges over that age do not. Years on the bench is related to age, but this factor does not clearly divide the two groups of judges. Although all five of the new members of the court participate, they are joined by three of the more experienced judges, two of whom have been on the bench for ten years. The council thus allows the newer judges to

become familiar with the court's sentencing practices.⁵⁰

There are two criteria that distinguish the cases the participating judges withhold from the council from those they bring before it: the seriousness of the offense and whether the conviction was handed down by a jury. Judges are most likely to bring before the council the more serious criminal cases and those cases in which conviction was obtained by jury trial (Table 25).

Table 25

PROPORTION OF CASES BROUGHT TO THE COUNCIL
BY SEVERITY OF THE CRIME* AND TYPE OF CONVICTION
(PERCENTAGE ARE BASED ON THE NUMBER OF CASES IN EACH CELL)

	No Custody	Custody
Guilty Plea	35% (573)	60% (318)
Bench Trial	42% (19)	61% (37)
Jury Trial	83% (18)	73% (84)

*as measured by the sentence actually imposed for noncouncil cases and by the sentencing judge's initial sentence recommendation in council cases.

Sixty percent of all guilty plea cases likely to result in prison sentences were brought to the council; only 35 percent of the non-custody guilty plea cases were brought to the council.⁵¹ Possibly the judge considers the duration of a prison sentence to be a more important issue than the length of probation and therefore is more interested in his colleagues' advice.⁵² Another reason may be that there is greater variation in the length of custody sentences. These percentages hardly change if the conviction was obtained after a bench trial. But if there was a jury trial, the proportion of cases brought to the council was high, regardless of whether the case was likely to end in a custody sentence.

Finally, the voluntary character of the Chicago council weakens its power to reduce overall sentence disparity in the court, because

⁵⁰ Two of the three more experienced participating judges bring few of their cases before the council, thus reinforcing the notion that the council serves an educative function.

⁵¹ The classification custody-noncustody is based upon the initial recommendation of the sentencing judge.

⁵² We have not determined for this article the types of cases—distinguished by crime, defendant, and mode of conviction—in which the sentencing judge is more (or less) likely to accept the advice of the council.

only one-third of all cases come before the council. To estimate the council's effect on the entire court, it is necessary to estimate the disparity in the two-thirds of all cases not brought before the council.

The participating judges fail to bring one-half of their cases before the council; this represents one-third of the total cases. There is evidence that these cases tend to be low-sentence, noncustody cases,⁵³ which have a somewhat lower average disparity.⁵⁴ On the other hand, the one-third of all cases heard by nonparticipating judges will tend to involve above-average disparity: four of the six nonparticipating judges are found at the extremes of the severity spectrum; one near the lower extreme, two near the upper extreme, and one judge so severe that he exceeded the upper extreme.⁵⁵

If we assume that these two deviations offset each other, we may conclude that the two-thirds of all cases that do not come before the council have the same average disparity as the one-third that come before it. A 10.7 percent reduction in disparity was found for the one-third of the Chicago cases brought before the council. Assuming that the remaining two-thirds have the same average disparity, this 10.7 percent decrease in disparity must be reduced by two-thirds to reflect the council's effect on all cases sentenced in the Chicago court. Therefore, the council reduces the sentence disparity of all cases in the Chicago court by not quite 4 percent.

VIII. THE NONCOUNCIL-RELATED CHANGES

At an earlier stage of our analysis we decided temporarily to disregard changes by the sentencing judge that could not be meaningfully related to the council's deliberations. We tentatively assumed these changes were induced by events that occurred after the council finished its deliberation. By definition, all of these changes were away from the mean, and hence would have increased

⁵³ See Table 25 *supra*.

⁵⁴ The New York court shows the same pattern:

Chicago	
Mean Severity	Percent Disparity
.01 - 3.49	29.8
3.5 - 6.99	46.7
7.0 or higher	36.4

⁵⁵ This ranking was derived from the net percentage of cases in which a judge imposed custody if the Probation Department recommended probation, minus the cases where the judge gave probation where the Probation Department recommended custody. This was the only severity measure for which we had data on all judges, including those who did not participate in the council.

disparity. They, too, constitute a major difference between the two courts: in Chicago, their number was negligible, between 6 and 8 percent of all cases, depending on the mode of computation; in New York, their number was between 15 and 22 percent.⁵⁶ The judges in New York explained that these changes, all sentence reductions, were made on the basis of the information conveyed in the sentencing hearing. There the prosecutor, as a rule, stands silent and only the defense speaks. Although all of these moves were in the direction away from the counseling judges' recommendations, we decided not to consider them as increasing disparity, under the optimistic assumption that the counseling judges would have changed with the sentencing judge had they become aware of the new information. But, since only New York shows this high frequency of unexplained reductions, the puzzle remains. We hope to resolve it eventually by tracing these changes and having them explained as they occur.

IX. DISCUSSION AND CONCLUSIONS

We have analyzed many facets of the sentencing councils, and we shall now summarize our findings and discuss their limitations. We shall also propose possible alternatives a court might adopt to help reduce sentence disparity.

Here, first, is the summary of our findings—the amount of sentence disparity that exists among the judges of the two courts, the amount removed by the sentencing council, and the amount of disparity that remains (Table 26).⁵⁷

In each court the council is able to reduce about 10 percent of the sentence disparity in the cases that come before it. In Chicago, since only one-third of the cases are brought before the council, the reduction in all cases is under 4 percent.

To evaluate this statistic, one must see its limitations. First, the roughness of the measuring scale fails to detect such minor but potentially important nuances as special treatments or conditions accompanying a sentence.⁵⁸

⁵⁶ See text at note 44.

⁵⁷ See Table 18 *supra*.

⁵⁸ In our one observation of a council meeting in Chicago, we could see several instances of this role of the council. In one case, discussion turned on whether the offender should be sentenced under the provisions of the Youth Corrections Act which allows the conviction record to be expunged. In another case, the judges discussed the feasibility of a work-release form of incarceration which would allow the offender to keep his job. A third discussion hinged on the probation officer's recommendation that the offender be required to move his residence in order to avoid the temptation of further gang activities.

Table 26
 EFFECT OF COUNCIL ON DISPARITY
 (LIKELY DIFFERENCE IN THE SENTENCES
 OF TWO JUDGES SELECTED RANDOMLY FROM THE COURT)*

	Cases before the Council	Chicago	New York
		All Cases	All Cases
(a) Disparity Prior to Council Deliberation*	36.7	36.7	45.5
(b) Percent of Disparity Removed by the Council*	10.6	3.6**	9.6
(b x a) Percentage points removed by Council	3.9	1.3**	4.4
(a-b x a) Disparity after Council Reduction*	32.8	35.4	41.1

*From Table 18

**Because of voluntary participation, only 1/3 of all cases

More importantly, the study observes reduction only within the narrow time interval between discussion in the council and the imposition of sentence. It thus does not indicate anything about the possible long-range effect of the council. It may be that the very existence of the council tends to reduce disparity and that the disparity observed among initial recommendations would be greater were it not for the experience gained by the judges in previous meetings.

In order to determine whether the council has this long-range effect, it would be necessary to measure its disparity at different points in time and to compare these data with parallel measurements in a court with no council. We retrieved rudimentary data on such a comparison from the study of the Second Circuit by the Federal Judicial Center.⁵⁹ That study asked each district judge in the Second Circuit for his sentence recommendations on an array of twenty selected cases. By comparing the judges of the Eastern District (who participate in a sentencing council) with those of the Southern District (who do not), we were able to compute the sentence disparities for the two districts (Table 27).

The disparity in the Eastern District is not smaller, and is even somewhat larger, than in the Southern District. Table 27 provides an important by-product. The disparity figure computed for the

⁵⁹ See note 26 *supra* (our computation).

Table 27

SENTENCE DISPARITY IN A COURT WITH A
SENTENCING COUNCIL AND IN A COURT WITHOUT ONE

	Eastern District of New York (Before Discussion by Council)	Southern District of New York
From the Federal Judicial Center Study of Simulated Cases	47.8%	42.0%
From the Sentencing Council Data	45.5%	

Eastern District from the Federal Judicial Center study is nearly identical to the disparity measure derived from the sentencing council, suggesting that simulation results may parallel sentencing council results.

The data are only suggestive, since the comparison refers to only one point in time and we do not know the disparity with which each court began. But to sustain the notion of a long-range effect of the council in reducing disparity, we would have to assume that the original sentence disparity in the Eastern District was even larger than the observed 47.8 percent. This is a difficult notion to sustain since the two districts are adjacent and similar and their judges come from the same reservoir.

There are inherent difficulties in measuring reduction in disparity over time. The sentencing practices of individual judges may change, and the composition of the court will change as judges retire and are replaced. To assess the full impact of the council on this learning process, it would be necessary to trace the sentencing patterns of individual judges over time.

Again, a substitute measure was developed by grouping the judges according to their years on the bench and then comparing their average percent distance from the mean, as in Table 28.

TABLE 28

AVERAGE PERCENT DISTANCES FROM THE MEAN FOR
JUDGES* GROUPED BY YEARS ON THE BENCH

	1 year to 2 years	Over 2 years to 5 years	Over 5 years
Chicago	26%		26%
New York	29%	26%	27%

*for all cases in which judge gave a recommended sentence

Although this table does not indicate the disparity of the more experienced judges when they first joined the bench, the near uniformity of disparity levels suggests that further research may not disclose that disparity significantly declines over time in a court with a sentencing council.

Finally, the present study was primarily actuarial, comparing the sentences recommended before the council convened with the sentences eventually imposed. We have observed no council meeting in New York and only one in Chicago. Council meetings may be useful simply as a relief from the isolation in which the judge finds himself. One participant in the sentencing council of the United States District Court for the Southern District of Michigan described the process with sensitivity:

It is here, in the give and take of fraternal discussion, that points are emphasized or subordinated according to the judgment of the individual judges, with the result that there is a close approach to a common meeting ground. The weights assigned to the various factors thought to be controlling as to disposition of the case are sometimes modified by the sentencing judge in the light of the experience of his brother judges with their own previous sentences.⁶⁰

These qualifications notwithstanding, a simple balance remains: against an effort of some two to three hours per week for each judge, the sentencing council cures not more than 10 percent of the disparity in the cases that come before it. In New York, disparity is reduced from 45 to 41 percent. In Chicago, the original disparity of some 37 percent is reduced by 4 percent in the cases that come before the council, and thus only 1 percent across all cases.

There might be ways of improving the effectiveness of the council. One would be to require the sentencing judge to adopt the median⁶¹ of the recommended sentences. While such a rule would effectively reduce disparity, it might reduce the autonomy of the judges to an unacceptable degree.⁶² The ultimate goal of all arrangements

⁶⁰ Quoted in Smith, *supra* note 31, at 9.

⁶¹ The median is the point at which half of the recommended sentences are more severe, and half are less severe. Unlike the mean, it cannot be influenced by judges who try to increase their weight by submitting a recommendation more extreme than the sentence they would actually impose.

⁶² Sherwood Zimmerman, a student of Professor Leslie Wilkins at the State University of New York at Albany, simulated group sentencing decisions under three different procedures: (1) the group has purely advisory capacity; (2) consensus should be reached after deliberation; (3) the mean sentence of a secret ballot becomes the sentence of the court. He

to reduce disparity must not be to average the various sentence recommendations but to bring the initial recommendations closer together.

A promising way to achieve that goal might be to expand upon the information role performed by the council in acquainting the sentencing judge with the sentences his colleagues would impose. Courts could set up internal reporting systems of all sentences imposed by their judges or the judges of their circuit. Already used in many courts, computers could provide the judges with the distribution of sentences, together with their means and medians for any combination of crime and offender. Eventually, such data could form the foundation for meaningful sentencing guidelines, especially if they included information about the parole board decisions and hence the relationship between sentence and actual time served.⁶³ One might consider developing such an information system either as a supplement to the council or as a substitute for it. This approach should be particularly helpful in nonmetropolitan districts, where judges, because of the geographic isolation, are unable to attend council meetings.

At this time, the most determined efforts to reduce disparity through clear but flexible guidelines are made at the last stage of the regular sentencing process by the United States Parole Board.⁶⁴ The sentencing council, even if its reduction effect is small, represents one of the very few institutional attempts to resolve the disparity problem. The participating judges, although aware of its modest impact, like the council and find the time they devote to it well spent.

found procedure (1), modeled after the sentencing council, least likely to reduce disparity within and between groups and procedure (3) most likely to reduce it; procedure (2) leads to intermediate results.

⁶³ The multiplicity of sentencing choices available to the courts, and the varying attitudes between sentencing judges results in a wide disparity in the lengths of sentence imposed on persons convicted of similar offences and often who possess similar backgrounds. To a very real degree, the Board of Parole tends, in practice, to equalize this disparity whenever it is not bound to the one-third maximum line required in regular sentences.

UNITED STATES BOARD OF PAROLE, BIENNIAL REPORT 1968-1970, at 13 (1970).

⁶⁴ Through bold integration of research and policy making, Professor Wilkins, *supra* note 62, under another grant from the Law Enforcement Assistance Administration, has joined the U.S. Parole Board in a broad reform move. The examiners and the board are now operating under clear guidelines, which take into consideration the gravity of the committed crime and the record of the offender. The prisoner's behavior in the institution allows the examiner to deviate slightly from the guidelines. If he cares to exceed the limits, he needs approval of the board. Unless, therefore, the sentencing judge prescribed a minimum sentence, the parole board imposes in fact a *de novo* sentence. This is an interesting reflection on the sentencing system and might well hasten its reform. Some of the underlying thinking is reported in W. AMOS & C. NEWMAN, *PAROLE* (1975).

We cannot disagree. In the absence of legislative reform efforts,⁶⁵ the sentencing council is a progressive innovation that deserves support, if only as the first step in the resolution of a troublesome problem. An average sentence disparity between any two judges of around 40 percent is too much.

⁶⁵ Some state courts have attempted to reduce sentence disparity through sentence review by a panel of three trial judges, which may not include the sentencing judge, *after* sentence has been pronounced.