

**The Role of Gender in Determining the Criminal Sanction:
Results from Multimedia Experiments in Criminal Sentencing**



Jerry Goldman
Department of Political Science
Northwestern University
Evanston, IL 60208
j-goldman@nwu.edu



Kent E. Portney
Department of Political Science
Tufts University
Medford, MA 02155
kportney@tufts.edu

Prepared for presentation at the Annual Meetings of the American Political
Science Association, August 28 - 31, 1997, Washington, D.C.

No part may be reproduced without explicit permission from the authors.
© 1997 Jerry Goldman and Kent E. Portney

Acknowledgements

We would like to thank Prof. Wesley Skogan of Northwestern University for his cooperation in testing one round of experiments in client/server mode with his students. We would also like to thank Giel Stein, Da-Wai Hu, and Erik Connelly for their research assistance. We are grateful to Dr. Steve Cohen, Director of the Tufts Curricular Software Studio, for preparing the data and for comments and criticisms on earlier drafts. Finally, these experiments would never have come to pass without the skill and wizardry of Mitchell Rosenbaum of Grafix Hypderdot Media, Inc. We would also like to thank the over 200 students at Tufts and Northwestern for agreeing to participate in the experiments on criminal sentencing.

**The Role of Gender in Determining the Criminal Sanction:
Results from Multimedia Experiments in Criminal Sentencing**

Jerry Goldman
Northwestern University

and

Kent E. Portney
Tufts University

This paper addresses the role of gender in the criminal sentencing process. Studying gender effects presents thorny substantive and methodological challenges in research on sentencing disparities -- how one goes about discerning whether, to what extent, or under what conditions “gender” or other “extra-legal” factors impact the punishments meted out by judges.

The paper has six parts: first we discuss and create the context of the general problem of sentencing disparity; second, we discuss the specific issue of the role of gender in determining criminal sentencing within this context; third, we address the methodological problem associated with the study of sentencing behavior, particularly with trying to isolate the effect of one variable compared to others in determining sentence severity; fourth, we examine one alternative methodology that has rarely been used -- the experimental approach; fifth, we present the results from analysis we have conducted using an innovative experimental approach; and finally, we discuss the implications of our findings.

I. The General Context of the Sentencing Disparity Problem

Judges exercise power when they sentence criminal defendants. Prior to the implementation of sentencing guidelines (now in use in federal courts and at least seven states), judges possessed extraordinary discretion in sentencing. This discretion led to concerns that sentences were frequently a function of extra-legal considerations made by a judge and other parties involved in sentencing when imposing a punishment rather than a function of the more legitimate characteristics of the crimes committed and relevant attributes of the defendant. (Partridge and Eldridge, 1974; Neubauer, 1988) With sentencing guidelines, reforms sought to constrain, if not eliminate, the influence of illegitimate factors by narrowing the range of permissible sentences. Today, after a number of years of experience with the implementation of sentencing guidelines, new reform advocates make a compelling case for the wise use of greater sentencing discretion. They argue for more flexibility in sentencing to avoid what they consider manifest injustices resulting from the tight control on sentencing imposed by legislatures or sentencing commissions. Many other reforms have been proposed that would effectively constrain, albeit less directly, judge discretion, including President Clinton's recent call for a Constitutional amendment to ensure that crime victims and

their survivors would play a significant role in the judicial process, including sentencing. (Sandel, 1997)

Sentencing guidelines emerged from a decade of research conducted during the 1970s. With or without sentencing guidelines, any amount of discretion in sentencing raises the possibility that extra-legal factors could influence judicial behavior. As we note later, empirical research on this issue remains divided. Yet the issue is as alive today as it was during the high water mark of studies of sentencing discretion and its results. In a recent, much publicized, study in New York state, for example, researchers purported to find that African American and Hispanic defendants were treated considerably more harshly than their white counterparts (Nelson, 1995). A core issue concerns the manner in which researchers study sentencing issues and their ability to isolate effects in their causal models.

We offer a brief review of the empirical research on disparity and the role of extra-legal factors such as race or gender. Our purpose is to provide a sense of the findings from, and difficulties associated with, the prevailing approaches to studying sentencing disparity.

Disparity in the sentencing process within the criminal justice system of the United States has long been noted as pervasive and endemic. The most widely accepted and used definition of sentencing disparity is provided by the U.S. Congress: "...disparity exists when defendants with similar criminal records found guilty of similar criminal conduct receive dissimilar sentences." (28 U.S.C. Section 991 (b)(1)(B)) Yet even with this definition, whether there exists sentencing disparity has been the subject of decades-long debates.

The Second Circuit Sentencing Study (Partridge and Eldridge, 1974) stands out as a model of careful experimentation merged with simulation. The purpose of this disparity study was to observe the sentences of many judges in identical cases. Fifty judges rendered sentences on twenty cases selected to represent the business of the federal trial court that comprises the Second Circuit (southern New York, Connecticut, and Vermont). These same judges also rendered sentences on ten other cases selected to test for case characteristics that might explain any disparity. Here the investigators randomly varied relevant (e.g., prior record) and purported irrelevant (e.g., race) factors in explaining the sentencing behavior of their subjects.

In the first part of the study, the investigators found substantial disparity in the sentences across the twenty cases. There were large differences in prison terms imposed in the same case. In most cases, the judges lacked agreement on whether any incarceration was appropriate. In the second part, the researchers tested the effect of

probation office recommendation, heroin addiction, method of conviction, prior record, and socio-economic background considerations. In only one hypothesis (prior criminal convictions) did they find a predicted effect on sentence length. In the presence of a presentence report containing evidence of a prior criminal record, judges were inclined to sentence such a defendant more harshly than in the absence of such evidence.

Following Partridge and Eldridge's sentencing study of the Second Circuit, at least one simulation and several nonexperimental studies have been conducted to determine the existence, if any, of disparity in the sentencing process.

Austin and Williams conducted a paper survey in 1975 in which forty-seven Virginia district court judges were given an identical booklet containing descriptions of five legal cases. None of the independent variables of the defendants were altered, and guilt was not assumed. The judges were to recommend a verdict, and, if a guilty verdict was rendered, to suggest an appropriate sentence based on Virginia law. Their purpose was to gain a relatively "pure" estimate of the degree of sentencing disparity among the same type of judges within the same state jurisdiction. They found that judges showed statistically significant and substantial disparities; that the strength of the evidence against a defendant is related to disparity; that strong evidence does not appear to guarantee a high agreement on the appropriate sentence; that the type of offense appears to be related to the degree of disparity; that the five cases produced a variety of patterns of disparity, but some form of disparity was always present; and that in their data, the rate of agreement on verdicts varied from 38.3% guilty verdicts to 100%, preference for a single mode of sentencing varied from 29.3% to 83.9%, and the magnitude of sentences within modes was absolutely large for the minor offenses studies. In their words, "these figures, though roughly calibrated, are evidence of 'substantial' disparity" (Austin and Williams, 1977: 310).

McDonald and Carlson (1993) examined racial and ethnic disparities in sentences imposed on Federal offenders before and after implementation of the Sentencing Guidelines and the Anti-Drug Abuse Act of 1986. They analyzed patterns in sentences for whites, blacks, and Hispanics by controlling for explanatory variables that may be correlated with race or ethnicity. Their approach differed from other studies in that they did not take the guideline range as the standard against which sentences should be evaluated. Rather, they examined the actual sentences imposed, and they considered the guideline range as one constraint among many that may affect the sentencing decision. Moreover, because the guideline ranges can still be rather large, substantial variation in racial differences may persist even within guideline compliance. Thus, McDonald and Carlson examined the differences in actual sentences given to offenders, regardless of the guideline ranges. The approach in general is to attempt to

account for differences in sentences between races by statistically holding constant the other differences among whites, blacks, and Hispanics that may explain the differences in sentences.

Sentences in the McDonald-Carlson study were characterized by two dependent variables: whether the defendant was incarcerated, and for those who were, a measure of the length of prison sentence imposed. Prior record variables were also correlated with both race and sentence severity. In every case, the goal was to compare equivalent groups of black and white offenders, based on variables related both to sentence and to ethnicity.

The investigators found: 1) During 1986-1989, before full implementation of the guidelines, white, black, and Hispanic offenders received similar sentences, on average, in Federal district courts. 2) Between 1989 and 1990, there were substantial differences in sentences for white, black, and Hispanic offenders. Specifically, 85% of Hispanic offenders and 78% of black offenders were sentenced to imprisonment, compared with 72% of white offenders; black offenders sentenced to prison during this period had imposed sentences that were 41% longer than for whites. For incarcerated Hispanics, the average imposed sentence did not differ significantly from whites. As McDonald and Carlson note, however, "[n]early all of the aggregate differences...can be attributed to characteristics of offenses and offenders that current law and sentencing guidelines establish as legitimate considerations in sentencing decisions" (1).

The United States Sentencing Commission carried out its own study on disparity in 1991. To address the issue, the Commission examined sentences imposed for offenders convicted of bank robbery, embezzlement, heroin, and cocaine offenses, matched for similar offense and offender characteristics. As their reasoning stated, "these categories were selected to ensure adequate samples at the aggregate level and to examine offense types that represent a varied cross-section of federal crimes (i.e. street crimes, white-collar crimes, and drug offenses) that make up a large proportion of the federal caseload" (U.S. Sentencing Commission, 1991: 33). The Commission examined length of incarceration for similar offenders convicted of similar offenses both in pre-guideline years and under the guidelines. The Commission found that within each case, disparity had been reduced considerably under the guidelines by narrowing the range of sentences that had previously been imposed; in all, most sentences became more similar in a statistically significant manner.

The Federal Judicial Center (1981) issued a study measuring the effects of sentencing councils on sentencing disparity in the federal courts in Chicago, Detroit, and Brooklyn. The study concluded that "the effect of introducing a sentencing council into a court cannot be predicted without knowledge of many court

characteristics, so no conclusion can be drawn regarding the general usefulness of the councils in reducing disparity" (Federal Judicial Center, 1981: 14). Their data indicated that each district court council had reduced disparity, increased it, and in some cases had not affected it at all. Thus, while recognizing that their data sample was small, they found that they could not predict how sentencing councils would affect disparity in sentencing.

Diamond and Zeisel (1975) conducted their own study of how sentencing councils affected disparity in New York and Chicago. Using statistical tests on the data, they concluded that in each court, the council removed about 10 percent of the existing disparity, thus reducing disparity in New York from 45 to 41 percent, and in Chicago from 37 to 33 percent.

A 1988 RAND Corporation study based on over 10,000 pieces of data from California's Board of Prison Terms was used to determine whether racial bias in sentencing persisted after the 1977 Determinate Sentencing Act. (Klein, Turner, and Petersilia, 1988) They concluded that in California, defendant race is not related to the sentence imposed, once there are controls for relevant crime, prior record, and process variables. This finding appeared robust. It held for the decision whether or not to sentence the offender to prison or to probation as well as applying to the length of sentence imposed. Thus, the Act did appear to contribute to racial equity in sentencing.

Joan Petersilia (1983) conducted a study examining racial disparities in the criminal justice system for the Department of Justice. The study data was obtained from the California Offender-Based Transaction Statistics and the RAND Inmate Survey, which contains data on 1,400 inmates in California, Texas and Michigan. Using multiple regression analysis of system decisions and criminal behavior to control for major variables that might create the appearance of discrimination, Petersilia was able to conclude that minorities receive harsher sentences and serve longer prison terms. In California, regression analysis showed that length of sentence was significantly related to age, offense type, race, prison infractions, and marital status. Specifically, blacks serve 2.4 months longer than whites, and Hispanics serve 5 months longer than whites. In Texas, blacks serve sentences that are 7.7 months longer, and Hispanics serve sentences that are 8.1 months longer than whites'. In Michigan, time served depended primarily on age, offense type, juvenile record, and time spent in segregation. Race had no statistically significant effect on sentence length served.

Hewitt (1975) studied the effects of individual characteristics on judicial sentencing. Data for the analyses employed were obtained from 504 convicted adult felons in Seattle. The eight individual resource variables considered are: sex, age, race, marital status, education, socioeconomic status, work history, and dependents. Using

causal modeling and path analysis techniques, Hewitt concluded that only five of the eight individual resource variables had statistically significant total effects. Those five were sex, race, education, work history, and dependents. Of those five, only the sex of the defendant and whether there were dependents had significant direct effects on any of the sentence variables. Thus, unwarranted sentencing disparity could be seen to exist in this sample. Specifically, while only half the males receive deferred sentences, nearly 80 percent of the females do. Also, while 63 percent of these defendants received deferred sentences, only 37 percent without dependents received such sentences. And only 10 percent of those with dependents received incarceration sentences, but 27 percent of those without dependents received this sentence. Finally, Hewitt also found that there was no direct effect of race on any of the sentence variables. The total effect of race appeared to be largely mediated by the work history, the offense related variables, and bail. However, as Hewitt also found, for incarceration sentences, there was a direct effect of race on the probation presentence recommendation of incarceration.

Blumstein et. al. (1983) of the National Research Council published a collection of essays on sentencing theory and practice. In it, Hagan and Bumiller (1983) reviewed some of the major studies on race and sentencing disparity. From their findings, they concluded that race was a statistically significant factor in contributing to sentencing disparity according to Clarke and Koch (1977), Gibson (1978), Hagan and Bernstein (1979), Kelly (1976), LaFree (1980), Pope (1975), Thomson and Zingraff (1981), and Unnever, Frazier, and Henretta (1980); however, Lizotte (1978) and Myers (1979) concluded that race did not significantly contribute to sentencing disparity in their respective locales of study.

The most recent disparity study by the New York State Division of Criminal Justice Services, Office of Justice Systems Analysis, (Nelson, 1995) examined case processing and outcomes for adults arrested for felony offenses between July 1, 1990 and June 30, 1992. After controlling for differences in variables that legitimately affect case processing decisions (e.g., prior criminal histories, seriousness of arrest charges, type of arrest charges, gender, county, and youthful offender eligibility), the analysis revealed that minorities, defined as blacks and Hispanics, were held in jail at indictment and sentenced to incarceration more often than comparably situated whites. Of the approximately 300,000 felony arrests in New York State over the two-year period, 14 percent of the minorities but only 9 percent of the whites were sentenced to prison. And 18 percent of the minorities but 12 percent of the whites were sentenced to jail. As the study authors stated, "[approximately one in three minorities sentenced to jail would have received a different sentence if they were processed as comparably situated whites. Disparities in sentencing decisions account for approximately three hundred prison and four thousand jail sentences of minority defendants each year." (Nelson, 1995)

Despite years of research on the role of race, the studies remain equivocal. Some investigators line up on the side of substantial race effects; others take the opposite view. And while both views may be correct given the boundaries of any research endeavor, no consensus has emerged on the impact of race.

II. The Role of Gender

In contrast to studies of race, empirical studies on the role of gender are rare. Women are a small fraction of all defendants, though in absolute but not relative terms, that number is increasing. And gender perspectives are relatively new to the social sciences, especially in relation to racial issues.

Daly's studies represent the most thoughtful and careful empirical work on the role of gender in sentencing. Her work is creative and provocative. We examine three of her recent studies to locate the state of knowledge with regard to gender and sentencing and to identify what we see as a methodological weakness at the core of virtually all recent gender- and race-based investigations.

Study 1. In *Gender, Crime, and Punishment* (1994), Kathleen Daly conducted an analysis of gender effects using two distinctly different approaches. The study focuses on a "wide sample" of all 189 women and 208 randomly selected men sentenced for felony convictions in New Haven Superior Court between 1981 and 1986. In the first approach, she examined the magnitude of gender differences, and then subjected these cases to standard multivariate analyses to estimate the gender effects on sentence severity (incarceration vs. non-incarceration). In the second approach, she examined a "deep sample" which attempted to create a quasi-experimental design by identifying 40 matched pairs (one male, one female) of defendants.

On its face, there appeared to be significant gender bias in sentencing. In general, Daly found that the gender gap in incarceration rates was 29% (with men more likely to receive a prison sentence). Among people sentenced to prison, men received sentences that, on average, were 13.3 months greater than women (Daly, 1994: 25). Analysis of the wide sample using multiple regression indicated that the gender effect on incarceration and on length of incarceration was statistically significant even after controlling for variation in crime severity, prior record, race, ethnicity, having a public defender, the judge involved, and other variables.

Going beyond the multivariate approach, Daly selected a smaller set of cases for more intensive analysis.

Using a strict decision protocol and not knowing the sentencing outcomes, I selected forty pairs of men and women who were accused and convicted of the same (or nearly the same) statutory offenses. [This deep sample] contains an assortment of all the major crime types: interpersonal violence (including homicide and aggravated assault), robbery, larceny, and drug offenses. In selecting the deep-sample pairs, I used the traditional sociological notion of "equal treatment." I wanted to compare the crimes and punishments of men and women whose offenses were statutorily alike.... The results may surprise readers, as they did me. I find that for only one of the forty matched pairs was there a sentencing disparity I could not explain satisfactorily. The results of my pair-wise analysis stand at odds with the regression analysis, which shows a greater "gender gap" in sentencing. (Preface, no page numbers).

In summary, a standard multivariate analyses revealed gender effects on incarceration (dichotomized) or sentence length but the pair-wise effort controlling for all factors but for gender reveals no gender effects.

Study 2. Daly and Bordt (1995) analyzed fifty unique data sets from studies published through mid-1990. They found that out of the thirty-eight studies that controlled in some way for prior record, 26 percent found no gender effects, 29 percent found mixed effects, and 45 percent found significant effects favoring women. The study suggests that the kind of statistical technique employed in the analysis matters. In general, multivariate analyses that controlled for a variety of variables, especially prior record, attenuated findings of gender differences. But even when such controls were introduced, gender effects were common. Gender effects were more pronounced when the dependent variable was dichotomous (prison-no prison) than when the dependent variable was measured as the length of prison sentence.

Study 3. Daly and Tonry (1997) conducted an extensive review of numerous studies of the effects of gender and race on sentencing. They start with a comprehensive review of the fact that gender differences, on their face, are much greater than race differences. The study also reviewed analysis with regard to a gender and a race interaction. Studies generally find gender differences within race groups but not racial or ethnic differences within gender groups. From the small number of studies available, it appears that the gender gap may be widest or more often statistically significant for black defendants. (Daly and Tonry, 1997: 229-230)

The core concerns in these studies center on the ability to "determine whether men and women prosecuted for *like* crimes are punished differently" (Daly, 1994: 4 (italics added)). Daly is skeptical of conclusions of gender disparity resting solely on statistical (i.e., multivariate) analyses (Daly, 1994: 268). "If [such] a study finds gender differences, apparently favoring women, the authors should have the additional burden of demonstrating, by a case

analysis or other means, that the statistical result is correct.” (Surely this rule would apply whatever the findings, because accepting or failing to accept one’s research hypothesis rests on the same body of evidence.) But this only points out the central weakness of Daly’s study and others: the ability to locate and compare *like* cases. As we see the problem, the issue can best be addressed by determining whether or not men and women prosecuted for *identical* cases are punished differently. In the next section, we elaborate this problem and offer a design-based solution.

There is one additional limitation with existing research that should be mentioned. Although this has been the explicit subject of research in very few studies, the issue of the gender of the sentencing judge may well constitute an important part of understanding the role of gender in the criminal process. The fact that so few of the studies of gender effects on sentencing has examined the effect of the gender of the judge strongly suggests that the reason for this has more to do with the lack of data, and the rarity of female judges, than it does about the importance of the subject. For example, in Daly’s analysis in Study 1 discussed above, none of the judges in her study was female! The question concerning the effect of the gender of the judge on possible disparities in the sentencing of male and female defendants is nonetheless an important one.

In one of the few studies to examine the gender of the judge, Kritzer and Uhlman (1977) use data from over 22,000 cases disposed of in the courts of an anonymous major urban area between 1968 and 1974 to examine the relationship between gender of the participants (particularly the defendant and the judge) and sentences by various offense categories (manslaughter (n=433), robbery (n=5,250), aggravated assault (n=2,960), minor assault (n=2,678), larceny (n=3,890), forgery (n=1,064), drug offenses (n=6,187), and rape (n=886)). Female judges presided over 4.5% of the cases heard. Kritzer and Uhlman concluded that the female judges behaved no differently than their male counterparts, with both female and male judges in the group consistently sentencing female defendants less severely than male defendants who committed comparable crimes. Kritzer and Uhlman explained their findings along two lines of reasoning. First, they argued that the strong similarities in the career paths of all the judges studied had caused the female judges to become "masculinized" as they struggled to "make it in a man's world." This, combined with the possibility that female judges felt they had to overcompensate to prove themselves in the predominantly male legal world, diluted the effects of any early differences in sex role socialization experiences, resulting in a marked uniformity between male and female sentencing patterns. They speculated, however, that even though both male and female judges sentenced female defendants more leniently than male defendants, they did so for different reasons. While the lenience of male judges toward female defendants might have been considered a "chivalrous" display of masculinity, similar displays of leniency by

female judges toward female defendants were seen as the result of "sympathy" caused by the female judges' greater will and ability to identify with and understand the difficulties which drove their "sister" defendants to violate the law.

Second, Kritzer and Uhlman suggest that the similarity in male and female judge sentencing patterns may also be explained as the result of a universal "deep-seated" psychological prohibition against "inflicting physical pain on girls." Citing a study by Taylor and Epstein (1967), the authors claim that the prohibition against imposing serious criminal sanctions on females parallels the finding of a reluctance by research subjects to electrically shock misbehaving female subjects as opposed to equally troublesome male subjects. But men and women apparently exhibit this reluctance equally.

III. The Methodological Problem

With very few exceptions, virtually all studies have sought to examine sentencing disparities through the collection of data from actual cases in specific jurisdictions. The analysis of disparities typically proceeds with the bivariate examination of the relationship between the disparity variable (race, ethnicity, gender, etc.) and a measure of the severity of the sentence imposed. If men are found to receive, on average, more severe sentences than women, then a *prima facie* case can be made that disparities do exist.

Many studies seek to go beyond the bivariate analysis, recognizing that other facts could create spurious relationships with sentence severity. Indeed, almost all the empirical research on sentencing disparities has performed been rooted in some multivariate approach. The multivariate approach is motivated by the argument that one needs to find a way to separate the independent influence of gender or other disparity variable from other factors. The problem with bivariate analysis, as seen by the multivariate approach, is that gender could be related to other, perhaps more legitimate and legally relevant considerations, in which case presumably disparities either would not exist or would be considered of less concern. So, for example, if men are treated more harshly than women because they commit more serious offenses (as defined in the relevant criminal code), the multivariate approach suggests that this is much less bothersome than if gender disparities exist when the seriousness of the offense is "controlled."

Even if the spurious influence is no more legally relevant than gender, the issue of disparity can take on very different connotations. If, for example, one were to find that men are treated more harshly, but this pattern

disappears when the defendant's socio-economic status is "controlled," this carries with it a very different set of connotations than if indeed a pattern of gender bias persists even after socio-economic status is controlled. Thus, the research challenge becomes one of finding a way to control for *all* of the possible spurious influences on sentencing severity, a task which ultimately becomes unavoidably open-ended. The specification of just the additive hypotheses, and not accounting for possible interaction effects, using such an approach would then take the form:

$$Y_1 = \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \dots + \beta_n + e_1$$

where Y_1 is the dependent variable and $\beta_1 \dots \beta_n$ is the list of independent variables and e_1 is the error term for unexplained variance. Alas, the list of putative independent variables is limited only by the imagination of the researcher, theories of sentencing, or more pragmatic considerations such as whether a measure of a particular case characteristic is available to the researcher. Given the potential for a large (and conceivably, limitless) number of variables, the associated challenges of applying existing multivariate techniques to such data, and the often difficult task of convincingly interpreting the results of such analyses, raise serious questions about whether multivariate approaches can ever really provide compelling information concerning sentencing disparities. As a RAND study put it:

Our findings suggest that future correlational studies of racial disparity will need to control for a large number of factors. However, even with such controls, false conclusions can be reached, because race may be correlated with unmeasured but legally cognizable factors, such as the strength of the evidence presented. Because of this inherent weakness of correlational research, randomized experiments of pre-sentencing decisions under simulated but realistic conditions may help to identify the nature and size of racial biases at earlier stages of criminal justice processing. (Klein, Turner, and Petersilia, 1988: 12)

But the opposite is a problem as well: for a given sample of actual cases being analyzed, a factor that might be considered theoretically important may not vary in the sample. For example, Daly's study of criminal cases in New Haven yielded no cases in which the sentencing judge was female. Try as one might, multivariate analysis cannot create variation where there is none in the data.

Alternatively, however, field experiments in criminal processes are rare. Ethical and practical issues pose high barriers to such investigations. It is simply not possible for researchers to develop an experimental research design that would match defendants by race or gender and then randomly assign them to judges for sentencing. And

matching only approximates “likeness” of defendants when in fact ideal experiments call for *identical* defendant attributes but for the ones under scrutiny.

Another alternative, the use of paper-and-pencil-based simulated sentencing experiences with experimental designs, has proved useful even if rare in criminal justice. In the Second Circuit Sentencing Study we discussed earlier, researchers developed paper-and-pencil simulations in which they presented some fifty judges from the Second Circuit (NY) with twenty fabricated cases that were designed to be representative of the kinds of cases that are typically heard in that court. Although the study did not seek to examine racial or gender differences, it did uncover rather marked differences in sentences across judges.

The use of simulations allows researchers to experimentally control for case characteristics that otherwise might confound the isolation of sentencing disparities. For example, two cases could be devised that are identical in all respects except the gender of the defendant. Then, a randomly selected set of judges could be asked to sentence the man and the other half could sentence the woman. As long as the judges received the cases as a result of random assignment, differences in sentencing patterns may be reflective of disparities. In this controlled experiment, the underlying hypothesis would become much simplified compared to its earlier counterpart:

$$Y_1 = \beta_1 + e_1$$

where the only independent variable is the factor that has been varied in the experiment. Even if more than one variable is systematically varied, the resulting underlying hypothesis is considerably more tractable than its multivariate counterpart. Moreover, by *controlling* variation of each independent variable, alone and in combination with others, we can assess significance and estimate effects with economy and precision.

While the paper-and-pencil simulation overcomes many of the methodological problems found in other approaches (most notably the ability to control experimentally rather than statistically for case characteristics other than those of specific research interest), it is not problem-free. One weakness is that the paper-and-pencil simulation is devoid of the direct human element which may itself exercise influence on the sentencing process. Particularly when assessing the role of visually-relevant case characteristics, such as the race or gender of the defendant, one would expect this human element to take on great importance. After all, judges sentence human beings, not pieces of paper, though surely they rely on the latter in formulating their judgments. Certainly, cases presented purely on paper lack the verisimilitude that researchers might like to achieve in an experimental design.

The Second Circuit study remains a classic example of paper-and-pencil experimental research within the confines of the criminal justice system. The study is not without weaknesses, however. Perhaps the most critical weakness is the "paper defendant" problem. The judges gathered all their information about defendants from presentence investigation documents presented on paper. The absence of face-to-face contact removes a potentially important cognitive element usually found in actual sentencing processes. Judges' evaluations may stem from many sources, including the impression created by courtroom demeanor, eye contact, body language, and attractiveness. Direct contact also reinforces the passive information that may be found in a presentence report, such as race, gender, and age. There has been some research suggesting that such contact may convey important and subtle information to judges, and consequently produce differential impact.

Additionally, the logistical hurdles associated with implementing a paper-based experimental design simulation are non-trivial. The need to carefully manipulate paper cases in which some cases have one set of characteristics and other cases have some identical characteristics and some varied characteristics perform limits the number of variables that realistically and pragmatically can be varied, and ultimately limits the utility of a given piece of research.

IV. The Solution: "Crime and Punishment" Experiments

We have developed an alternative method that combines the power of experimentation with the verisimilitude of actual sentencing proceedings. We call the general approach the Applied Multimedia-Based Experimental Research (AMBER) approach. We have devised an application, called "Crime and Punishment,"¹ which permits us to implement an experimental design with multimedia, extending the basic utility of the paper-based simulation. This "Crime and Punishment" application calls for the research subject to sit at a computer and take on the role of a sentencing judge in sentencing some six cases which are composites of cases from Massachusetts and Cook County, Illinois.

The simulation contains an array of documents typically available to sentencing judges, and allows the

¹ A "Crime and Punishment" CD-ROM containing a non-experimental version of this multimedia simulation for either the MacOC or Windows 95 operating environments is available free of charge by writing to the authors.

research subject to take part in a sentencing hearing in which full-motion video presents the relevant actors - the prosecutor, defense lawyer, defendant, and the victim. The subject gets the “feeling” of walking into the courthouse, passing through the metal detector, acknowledging the greeting of court personnel, and entering his or her chambers. Once in chambers, the research subject picks up a court docket and folder and walks into the courtroom. Upon hearing "all rise," the subject proceeds to the judge's bench. The docket is placed on the desktop, and the research subject can select a case by clicking on any of six items on the docket. At this point, something that looks like a standard paper file opens up revealing two court documents describing the case at hand on one side, and a presentence investigation report on the other side. By clicking on either document, the research subject can read that document. When ready to do so, the subject can view a sentencing hearing by clicking on the gavel sitting on the desktop. The sentencing hearing consists of full-motion video of the prosecutor explaining the case and making a sentence recommendation; the defense lawyer presenting mitigating circumstances; the defendant making a brief statement; and the victim delivering an impact statement.

Crime and Punishment contains six different cases where the defendant in each case must be sentenced by each research subject. The first case encountered by the user, an armed robbery case, serves as an anchor and therefore was created with no variables, i.e. all research subjects see the exact same case. The remaining five cases, possession of drugs with intent to distribute, armed robbery, grand larceny shoplifting, and sexual assault on a minor (summarized in Table 1), are made accessible to the research subject through the court docket, as described above. Thus, the multimedia application is, from the outset, considerably more interactive than is its paper-and-pencil counterpart.

The utility of the AMBER system is readily seen in the way characteristics of the cases are manipulated. In the full implementation of “Crime and Punishment,” the researcher may vary the race of the defendant (the defendant could be either African-American or white); the gender of the defendant (either male or female); the appearance of the defendant (either wearing street clothes or in a prison jump suit); the affect exhibited by the defendant (the defendant might not speak, or when the defendant speaks, he/she can be either animated or subdued); and whether or not the victim makes a statement. By allowing only one of these characteristics to vary and holding everything else constant, we can isolate the effect of that variable on the sentence rendered by the subject. For example, we can vary the gender of the defendant and hold everything else constant. We do this by simply inserting a different video image into the courtroom scene while altering nothing else. Although we cannot adequately demonstrate the full power of this technique here on paper, in Figure 1 we present a sense of what the comparable images of two defendants, one male and one female, look like. The system allows the researcher to

vary more than one characteristic simultaneously, such as the race *and* the gender of the defendant. The data reported here are the result of just such an experiment where only race and gender were varied, while all other characteristics were held constant.

We conducted multimedia experiments with undergraduates at Tufts and Northwestern Universities. We acknowledge that there is a weak fit between 20-year-old undergraduates and much older elected or appointed law-trained judges. Our hope and expectation is that the experiments we report here will serve as prelude to research focused on subjects who are participants in actual criminal sentencing decisions - such as judges and prosecutors. We return to this issue in the next section.

We conducted our multimedia experiments with three groups of students: two groups of students at Tufts, and one group at Northwestern.² (We also conducted identical experiments using paper defendants as a control against the heretofore untested multimedia approach.) The three groups of students were virtually identical on virtually all characteristics. Given the strong similarity between the groups, we combined the data for our analysis. (Separate analyses did not produce between-group differences.) With the caveat discussed below, we allowed the gender and the race of the defendant to be randomly assigned to the students in each class, although we only report the results of the gender comparisons in this paper. We did not randomly assign case characteristics according to the gender of the student (judge) doing the sentencing. We simply recorded the student's gender, and used this as a variable in the analysis after the fact. The gender of the defendant was held constant across all subjects in two cases: the McKay armed robbery case; and the McNutt sexual assault case. In the former case, all students sentenced the exact same case as their first, anchor, case, i.e. all case characteristics were constants across all students (the defendant was a white male). In the latter case, we varied the race of the defendant, but to retain plausibility due to the nature of the offense, the gender of the defendant was always male.

In the multimedia experiments, students who acted as judges were provided with implicit information about the gender of the defendant. This information was provided in only two places in the entire simulated sentence process. In one of these, the gender of the defendant was designated as "F" or "M" in one small space, along with

² Students in Portney's "Judicial Politics" course took part during the spring semester 1996 and spring semester 1997; Students in Wes Skogan's "Introduction to Law in the Political Arena" course took part during the winter 1997 term.

the defendant's name, court docket number, and the age and race of the defendant, on the first page of the simulated paper "Court Document," which provides the judge with a summary of the facts in the case. In the other, the judge became aware of the defendant's gender by viewing the video-based courtroom hearing by viewing the defendant and listening to him or her speak. Our expectation is that the presentation of gender information in the experiment's simulation is contained in the same visual form as that found in actual criminal proceedings, and very different from that which is found in paper-and-pencil experiments which contain very different sets of gender cues.

The statements delivered in the video-based sentencing hearing were scripted from composites of actual cases in Middlesex County, MA and Cook County, IL. These statements - both the words that were uttered and the style in which they were delivered - were held constant across defendants. In other words, the statements made by the defendants were identical regardless of the gender of the defendant. As we discuss later, we are aware of the possibility that the scripted words are those of a male defendant and very different from those that might be spoken by a female defendant.

V. Findings

Our analysis focuses on the dual role of gender: does it seem to matter whether the defendant is male or female; and does the gender of the person doing the sentencing (the student-judge) seem to play a role. So the data we report are aimed at providing answers to these questions. Because the gender of the defendant was only varied in four of the six cases, most of our analysis concentrates on these cases.

There are many different ways of measuring the dependent sentence variable in sentencing studies, and we should say a few words about the measurement employed here. Although some sentencing studies rely on a dichotomous measure of sentence classifying defendants according to whether they received prison time or not, many others prefer to use an interval measure of the length of prison sentence or "time incarcerated," as Daly calls it. (Daly, 1994: 30-31) The analysis presented here is based on the latter, using the length of the prison term entered by the students where the length could theoretically range from zero months to the maximum number of months prescribed by statute. With this measure, we are able not only to distinguish those defendants who received prison from those who did not, we can also distinguish defendants according to the length of any prison sentence they received. Measurement issues create tough choices for sentencing researchers. There is no metric to calibrate sentences across their many modalities (imprisonment, probation - supervised and unsupervised - special conditions of probation, fines, and restitution). We explored the use of nonparametric approaches based on accepted rules of

sentence severity (any prison>any probation>any fine) but frequent ties in the sentences obscured meaningful analysis. Consequently, except where noted, we do not consider elements of the sentence other than the length of prison sentence.

Our analysis begins with an examination of average prison sentences. Figure 2 reports average prison sentences in the four felony cases by defendant's gender. Here we can see that female defendants receive consistently lower sentences, with the differences ranging from 1.1 months in the Canelly larceny case to 4.3 months in the Malloy armed robbery case. The direction of the difference is consistent with the view that female defendants are treated with greater deference, but the differences do not rise to the level of statistical significance. Figure 3 reports average prison sentences in the same four felony cases by the gender of the sentencing judge. Female judges meted out consistently lower sentences, with differences comparable to the data sorted by defendant gender. Still, none of the differences in means rises to the level of statistical significance.

We examine the evidence of judge gender from a slightly different perspective. In Figure 4, we view the average prison sentences by defendant gender within judge gender. Note that male and female judges mete out very similar sentences to male defendants. The more pronounced differences occur for female defendants. Female judges consistently appear more lenient when the defendants are females.

In Figure 5, we show the percent deviations from the overall means. Here we control for the magnitude of the sentence to account for the fact that small differences in lesser felonies (Caneely/larceny and Hart/restraining order) should be viewed in relation to the size of their means. In effect, we seek to normalize the sentences relative to their own case means. The results are pronounced. Female judges sentencing female defendants are always below the overall means. Indeed, the only other situation of a below average sentence arose when male judges sentenced the female defendant in the Hart case of domestic violence. Thus, men only appear to be willing to be lenient with a woman defendant who is engaged in domestic violence against a male. In contrast, male judges sentencing male defendants are always above the overall mean. The direction of these differences is greater in the female/female pairings in contrast to the male/male pairings. We should also note, however, that while female judges are more lenient on male defendants than their male counterparts, they still tend to mete out sentences to male defendants that are above the overall means.

Figure 6 displays the average prison and probation sentences in the four felony cases by judge gender within defendant gender. Our purpose is to see whether or not female judges have a tendency to substitute longer

probation periods (as an extension of the pursuit of rehabilitation) for prison sentences. Yet the average probation sentences appear to be very nearly constant across all subgroups. The one exception is when female judges sentenced female defendants in the Caneely/larceny case. In this case, the defense presentation relied on an argument that larceny of the type involved is a form of shoplifting, and this is akin to an addiction in which counseling and treatment are warranted. This argument seems to have convinced some judges to substitute probation with counseling for a prison sentence, but only when the defendant was female! Overall, however, it does not appear that female judges are substituting longer probationary periods for prison sentences.

What seems to account for these observed differences between female and male judges? One way of examining this is to look at the frequency distributions of prison sentences for each case, and to compare these distributions according to the gender of the defendant and judge. These frequency distributions are found in Figures 7 - 10. Each type of case produces a different frequency distribution, so we present the distributions for each type of case separately. Figure 7 shows the distributions for the Jameson drug case. In this way, we can look for the combination of gender characteristics that seems to yield a distinctive pattern of sentence distribution. When we examine these frequency distributions, we discover that female judge leniency is largely due to the fact that female judges rarely sentence either male or female defendants in the high end of the sentencing range. Nearly all of the highest sentences for each offense were meted out by male judges. For example, in Figure 8 showing the frequency distributions for the Malloy armed robbery case, we can see that the highest prison sentence given by a female judge was 84 months handed out by one judge; in contrast the highest prison sentence by a male judge was 132 months, and two judges handed this sentence out. Moreover, when female judges gave out lengthier prison sentences, they were more likely to be received by male rather than female defendants.

Finally, in Table 2, we report the analysis of variance for each case and for all four cases combined to identify the main effects of defendant gender and of judge gender, and of the interaction between the two variables. In the analysis of all cases combined, only the interaction effect is statistically significant. This is clearly a reflection of the tendency for female judges to sentence female defendants less harshly. In the Malloy armed robbery case, the gender of the judge produces a significant main effect, and again there is a significant interaction effect. For this armed robbery case, female judges are more lenient than male judges in general, and even accounting for this general leniency, they are still more lenient on female defendants. Beyond this, in the remaining three cases we cannot establish any statistically significant gender effects.

VI. Implications

We believe that the AMBER approach advances understanding of the complex mix of issues in criminal sentencing. The ability to control for gender (and other "sense-based" variables such as affect and attractiveness) offers the promise of new theoretical direction and new understanding to problems heretofore regarded as unresearchable. Counter to Daly, we see a consistent tendency in our data for the role of gender. Though we have not comprehensively established the statistical significance of gender effects, we cannot dismiss the fact that the direction of the data point consistently in a gendered fashion. And we have begun to identify the conditions under which gender does seem to make a difference. Social psychologists have long labored to explain human judgment from the myriad factors that influence it. From a theoretical perspective, we may profit from embracing the wide range of factors that impact human judgment in sentencing. In a digital world, our ability to use experimental controls is greatly enhanced. Such innovation should enlarge the range -- if not the direction -- of theoretical inquiry.

There are perhaps two main substantive implications which arise from the results we report here. First, efforts to champion the appointment and election of women to criminal courts promise to accomplish exactly what some of its proponents claim - to make the criminal justice system more sympathetic particularly to the plight of women. Clearly the results reported here suggest that women judges are more lenient with female defendants.

The second major implication relates to the impact of sentencing reforms on the effect of increasing the presence of females on the bench. As noted earlier, many states and the U.S. Congress have adopted sentencing guidelines or other mechanisms, such as mandatory or fixed-length sentences. The target of these reforms, by and large, has been judge discretion. Yet the fact is, limiting this judge discretion has the effect of diminishing the ability of judges - male and female alike - to respond to the needs of female defendants. To our knowledge, no set of sentencing guidelines calls for a different sentence to be imposed on a female versus a male defendant (Daly and Tonry, 1997: 206). In only one state - Minnesota - was the issue of the impact of sentencing guidelines on gender differences seriously considered, and the state opted for a gender-neutral set of guided sentences even if it meant that females would be treated more harshly than before. Analysis since the adoption of these guidelines confirms that sentences became less disparate by making sentences for female defendants more severe (Knapp, 1984). Similar patterns have been found in other states using sentencing guidelines as well (Daly and Tonry, 1997: 206). Thus, sentencing reform would restrict the discretion that would make it possible for female judges to treat female defendants more leniently. This strongly suggests that seating more female judges will not be able, by itself, to produce more lenient treatment of female defendants. Of course, whether this is an acceptable result is open to

debate.

We might also say a few words about the use of the multimedia approach versus more standard experimental approaches for the purpose of studying criminal sentencing. As a control against which to compare the multimedia approach, we conducted an identical set of exercises with students using a paper-and-pencil model. In these experiments, all case materials were presented to students on paper, and the assignment of versions of cases to students was done manually. All cues to gender were written in the court processing and presentence documents that subjects were asked to read. Our analysis was unable to discern any gender effects for defendants, judges, or interactions. This reinforces our view that applied multimedia is better able to introduce elements that are likely present in courtroom situation compared to analog methods of investigation.

The principal shortcoming of our study centers on our subject pool. Like most other studies in social sciences using experimental designs, this study's subjects are undergraduate students. Undergraduates at elite universities may be the stuff of future judges, but they certainly are not now judges. It would require a leap of faith we do not possess to claim that our data apply with the same force to judges as they do to undergraduates. Further testing with judges, judge-surrogates, or other parties typically involved in criminal sentencing will assure us that the leap of faith may be worth taking after all. Our intent is to move in that direction.

But even if we were to take this leap of faith, we recognize something of a shortcoming of our protocol as it relates to issues of gender. All of our video-based defendant statements were scripted based on actual cases in which the defendants were men. We essentially transposed the men's voices into female forms. Essentially, our experiments vary the female form and hold constant any gendered nature of the words that are spoken. Therefore, one might interpret our experiments as an inquiry into whether the gender form makes a difference when all else about the defendant is male constituted. But, of course, the question is would a woman have chosen to address the sentencing judge in the same way? In short, we have given the female defendants a male script. It remains an open question whether a woman's selection of words or phrases would have produced a different outcome.

Fortunately, we have a ready method to test this proposition in the future. In effect, this criticism boils down to whether or not a female-centered voice in a female form would cause a different response from judges, male and female. To test this proposition, we would identify women defendant statement's within the same case category (e.g., larceny or armed robbery, etc.), or at least devise scripts that are female-voiced, and then test the independent effects of script and gender (and their interaction) on subject response. This is not so far-fetched, as the

AMBER template will permit us to videotape, digitize, and test this hypothesis with relative ease. As we refine our conception of what gender means in the context of the criminal courts, we can modify our experiments to try to understand the role of gender in sentencing, and to sort out some of the many reasons why female defendants might be treated differently from male defendants (Leventhal and Krate, 1977; Sigall and Ostrove, 1975). We hope that this approach will be used by numerous researchers to replicate and improve upon the experimental protocols we have reported here.

REFERENCES

- Austin, William and Thomas A. Williams, III. (1977) "A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity," *The Journal of Criminal Law and Criminology*, Vol. 68, No.3, June, pp. 306-310.
- Blumstein, Alfred et al.. (1983) *Research on Sentencing: The Search for Reform, Vol. 2*. Washington, D.C.: National Academy Press.
- Clarke, Stevens H., and Gary G. Koch (1977) *Alaska Felony Sentencing Patterns: A Multivariate Statistical Analysis*. Anchorage: Alaska Judicial Council.
- Daly, Kathleen (1994) *Gender, Crime, and Punishment*. New Haven: Yale University Press.
- Daly, Kathleen and Rebecca Bordt (1995) "Sex Effects and Sentencing: A Review of the Statistical Literature," in *Justice Quarterly*, Vol. 12, No. 1, pp. 143-177.
- Daly, Kathleen, and Michael Tonry (1997) "Gender, Race, and Sentencing," in Michael Tonry, ed., *Crime and Justice: A Review of Research*. Chicago: University of Chicago Press, Vol. 22, pp. 201-252.
- Diamond, Shari Seidman and Hans Zeisel (1975) "Sentencing Councils: A Study of Sentence Disparity and its Reduction," *The University of Chicago Law Review*, Vol. 43, No. 1, Fall, pp. 109-149.
- Federal Judicial Center (1981) *The Effects of Sentencing Councils on Sentencing Disparity*. Washington, D.C.:

Federal Judicial Center, FJC-SP-81-2.

Gibson, James. L. (1978) "Race as a Determinant of Criminal Sentences: A Methodological Critique and a Case Study," *Law and Society Review*, Vol.12, pp. 455-478.

Hagan, John and Ilene Bernstein (1979) "Conflict in Context: The Sanctioning of Draft Resisters, 1963-76," *Social Problems*, Vol. 27, pp. 109-122.

Hagan, John and K. Bumiller (1983) "Making Sense of Sentencing: A Review and Critique of Sentencing Research," in A. Blumstein, et al., eds. *Research on Sentencing: The Search for Reform*, Vol. 2. Washington, D.C.: National Academy Press, pp. 1-54.

Hewitt, John David. (1975) *A Multivariate Analysis of Legal and Extralegal Factors in Judicial Sentencing Disparity*. Ann Arbor: University Microfilms International.

Hogarth, John (1971) *Sentencing as a Human Process*. Toronto: University of Toronto Press.

Kelly, Henry E. (1976) "Comparison of Defense Strategy and Race as Influences in Differential Sentencing," *Criminology*, Vol. 14, pp. 241-249.

Klein, Stephen P., Susan Turner, and Joan Petersilia (1988) *Racial Equity in Sentencing*. Santa Monica: RAND Corporation, R-3599-RC, February.

Knapp, Kay (1984) *The Impact of the Minnesota Sentencing Guidelines - Three Year Evaluation*. St. Paul, MN: Minnesota Sentencing Guidelines Commission.

Kritzer, Herbert and Thomas Uhlman (1977) "Sisterhood in the Courtroom: Sex of Judge and Defendant in Criminal Case Disposition," *Social Science Journal*, Vol. 14, pp. 77-88.

LaFree, Gary (1980) "The Effect of Sexual Stratification by Race on Official Reactions to Rape," *American Sociological Review*, Vol. 45, pp. 842-854.

Leventhal, Gloria and Ronald Krate (1977) "Physical Attractiveness and Severity of Sentencing," *Psychological Reports*, Vol. 40, February, pp. 315-318.

Lizotte, A.J. (1978) "Extra-Legal Factors in Chicago's Criminal Courts: Testing the Conflict Model of Criminal Justice," *Social Problems*, Vol. 25, pp. 564-580.

McDonald, Douglas C. and Kenneth E. Carlson (1993) *Sentencing in the Federal Courts: Does Race Matter? "The Transition to Sentencing Guidelines, 1986 - 1990."* Washington, D.C.: U.S. Dept of Justice, Office of Justice Programs, Bureau of Justice Statistics, NCJ-145328, December.

Myers, Martha A. (1979) "Offended Parties and Official Reactions: Victims and the Sentencing of Criminal Defendants," *Sociological Quarterly*, Vol. 20, pp. 529-540.

Nelson, James F. (1995) *Disparities in Processing Felony Arrests in New York State, 1990-1992.* Albany, NY: Office of Justice Systems Analysis, Division of Criminal Justice Services, New York State, September.

Neubauer, David W. (1988) *America's Courts and the Criminal Justice System*, 3rd edition. Pacific Grove, CA: Brooks/Cole.

Partridge, Anthony and William B. Eldridge (1974) *The Second Circuit Sentencing Study.* Washington D.C.: Federal Judicial Center.

Petersilia, Joan (1983) *Racial Disparities in the Criminal Justice System.* Santa Monica: RAND Corporation.

Pope, Carl E. (1975) *Sentencing of California Felony Offenders.* (National Criminal Justice Information and Statistical Service. Washington, DC: U.S. Department of Justice).

Ryan, John Paul, and James J. Alfini (1979) "Trial Judges' Participation in Plea Bargaining: An Empirical Perspective," in *Law and Society Review*, Vol. 13, Winter, pp. 479-507.

Sandel, Michael J. (1997) "The Hard Questions: Crying for Justice," *The New Republic*, July 7, p.25.

Sigall, Harold and Nancy Ostrove (1975) "Beautiful But Dangerous: Effects of Offender Attractiveness and Nature of the Crime on Juridic Judgment," *Journal of Personality and Social Psychology*, 31, pp. 410-414.

Taylor, S.P. and S. Epstein (1967) "Aggression As a Function of the Interaction of the Sex of the Aggressor and the Sex of the Victim," *Journal of Personality*, Vol.35, pp.474-96.

Thomson, Randall J., and Matthew T. Zingraff (1981) "Detecting Sentencing Disparity: Some Problems of Evidence," *American Journal of Sociology*, Vol. 86, pp. 869-880.

United States Sentencing Commission (1991) *The Federal Sentencing Guidelines: A Report on the Operation of the Guidelines System and Short-Term Impacts on Disparity in Sentencing, Use of Incarceration, and Prosecutorial Discretion and Plea Bargaining*. Executive Summary, Washington, D.C.: U.S. Sentencing Commission, December.

United State Sentencing Commission (1981) *Experimentation in the Law: Report of the Federal Judicial Center Advisory Committee on Experimentation in the Law*. Washington, D.C.: U.S. Sentencing Commission.

Unnever, James D., Charles Frazier, and Joan C. Henretta (1980) Race Differences in Criminal Sentencing. *Sociological Quarterly*, Vol. 21, pp. 197-207.

Table 1: Six Felony Cases Contained in the “Crime and Punishment” Multimedia Experiments

| Case Defendant | Felony Offense | Fact Summary | Was Gender of the Defendant Varied? |
|----------------|---|---|-------------------------------------|
| McKay | Armed Robbery | Defendant robbed a convenience store while masked; weapon was a knife | No |
| Jameson | Drug Possession with Intent to Distribute | Defendant possessed large quantity of powdered cocaine with intent to distribute; also possessed unlicensed handgun | Yes |
| Malloy | Armed Robbery | Defendant robbed an individual at knifepoint while victim was walking home from work | Yes |
| Hart | Assault, Violating a Restraining Order | Defendant committed domestic violence against live-in co-habitant; subsequently violated restraining order by breaking into house | Yes |
| Caneely | Grand Larceny | Defendant stole large amount of clothing from Bloomingdale’s Department store | Yes |
| McNutt | Sexual Assault | Defendant performed oral sex on the minor female daughter of his live-in girlfriend | No |

Table 2: Analysis of Variance Results Showing Gender Effects for Four Felony Cases

| Effect | Sum of Squares | df | F value | Significance level |
|---|----------------|-----|---------|--------------------|
| All Four Cases Combined | | | | |
| Defendant's Gender | 531.9 | 1 | 1.2 | .280 |
| Judge's Gender | 1563.9 | 1 | 3.4 | .064 |
| Defendant- Judge Gender Interaction | 3756.5 | 1 | 8.3 | .004 |
| Total Explained | 5907.7 | 3 | 4.3 | .005 |
| Residual | 363527.0 | 798 | | |
| Jameson Drug Case | | | | |
| Defendant's Gender | 251.5 | 1 | .482 | .489 |
| Judge's Gender | 340.2 | 1 | .651 | .421 |
| Defendant- Judge Gender Interaction | 487.5 | 1 | .933 | .335 |
| Total Explained | 1052.7 | 3 | .672 | .570 |
| Residual | 105004.3 | 201 | | |
| Malloy Armed Robbery Case | | | | |
| Defendant's Gender | 1077.3 | 1 | 2.9 | .089 |
| Judge's Gender | 1806.3 | 1 | 4.9 | .028 |
| Defendant- Judge Gender Interaction | 1426.7 | 1 | 3.87 | .050 |
| Total Explained | 4280.1 | 3 | 3.87 | .010 |
| Residual | 71564.5 | 194 | | |
| Hart Assault, Violating Restraining Order Case | | | | |
| Defendant's Gender | 189.7 | 1 | 1.64 | .203 |
| Judge's Gender | 24.6 | 1 | .212 | .646 |
| Defendant- Judge Gender Interaction | 4.1 | 1 | .035 | .851 |
| Total Explained | 233.3 | 3 | .670 | .571 |
| Residual | 22394.8 | 193 | | |
| Caneely Larceny Case | | | | |
| Defendant's Gender | 20.9 | 1 | .196 | .659 |
| Judge's Gender | 172.1 | 1 | 1.61 | .207 |
| Defendant- Judge Gender Interaction | 71.1 | 1 | .663 | .416 |
| Total Explained | 263.7 | 3 | .820 | .484 |
| Residual | 21236.4 | 198 | | |