Summer 1984

Types of Consensus in Public Evaluations of Crime: An Illustration of Strategies for Measuring Consensus

Terance D. Miethe

Follow this and additional works at: https://scholarlycommons.law.northwestern.edu/jclc

Part of the Criminal Law Commons, Criminology Commons, and the Criminology and Criminal Justice Commons

TYPES OF CONSENSUS IN PUBLIC EVALUATIONS OF CRIME: AN ILLUSTRATION OF STRATEGIES FOR MEASURING "CONSENSUS"*

TERANCE D. MIETHE**

I. INTRODUCTION

One of the most consistent findings of criminological research is that members of different social groups seem to agree on their ratings of the seriousness of crimes. This finding of consensus also appears to be extremely robust, with alternative scaling techniques, cross-cultural comparisons, and measures of related concepts (e.g., disapproval) yielding similar results.[1] Unfortunately, clear guidelines for establishing how much and what type of agreement qualifies as consensus have not been established. Moreover, even if various types of consensus exist in public evaluations of crime, an adequate explanation for these shared conceptions remains conspicuously absent.

This paper focuses on strategies for measuring "consensus" in public evaluations of crime. It identifies four distinct types of consensus on public ratings of crime in past research. To illustrate these types of consensus and explore the differences between them, various statistical techniques are used on a previously drawn sample of Baltimore residents[2] to measure the extent of each type of consensus. The types of consensus required for various applications of these studies to theoretical and policy issues are also discussed. The paper concludes with a discussion of

---

** Assistant Professor of Sociology at University of Wisconsin-Eau Claire.

[1] *See* G. NEWMAN, COMPARATIVE DEVIANCE: PERCEPTIONS IN LAW AND IN SIX CULTURES (1976); Miethe, *Public Consensus on Crime Seriousness: Normative Structure or Methodological Artifact*, 20 CRIMINOLOGY 515 (1982); Rossi & Henry, *Seriousness: A Measure for All Purposes?*, in HANDBOOK OF CRIMINAL JUSTICE EVALUATION 489 (M. Klein & J. Teilman eds. 1980).

[2] For a discussion of this sample, see Rossi, Waite, Bose, & Berk, *The Seriousness of Crimes: Normative Structure and Individual Differences*, 39 AM. SOC. REV. 224, 225-27 (1974).

the implications of this research for public policy and criminological theory.

II. OVERVIEW OF PAST STUDIES

Commentators have criticized studies of public evaluations of crime on a number of methodological grounds.[3] Researchers have addressed some of these criticisms, such as a sample selection bias[4] and the additivity assumption.[5] Other issues, such as the perceptual basis of seriousness ratings[6] and the extent of an aggregation bias,[7] remain and might severely restrict the utility of these studies as a basis for public policy. In fact, it is still unclear whether consensus in public ratings of crime truly exists and, if it does, what the reasons are for these shared evaluations. Definitive conclusions from past studies are not possible because researchers have failed to specify how much and what type of agreement is indicative of "consensus."

TYPES OF CONSENSUS AND STATISTICAL MEASURES

The dominant finding reported in past studies is that different social groups agree in their evaluations of crime.[8] Unfortunately, when consensus is reported as the major finding of a study, it is unclear what this means. There appears to be at least four types of consensus that can be obtained in public ratings of seriousness. These types of consensus are created by considering two issues: 1) the type of ratings (relative ordering or absolute magnitude), and 2) the degree of item inclusion (global or local inclusion). The four kinds of consensus and the statistical measures associated with each type are summarized in Table 1.[9]

---

[3] *See* Miethe, *supra* note 1, at 517-22.

[4] *See* Turner, *Introduction* to THE MEASUREMENT OF DELINQUENCY (T. Sellin & M. Wolfgang 2d ed. 1978) (discussing the reliance on college-student samples).

[5] *See* Wellford & Wiatrowski, *On the Measurement of Seriousness,* 66 J. CRIM. L. & CRIMINOLOGY 175 (1975) (discussing whether seriousness scores can be added to form complex events).

[6] *See* Miethe, *supra* note 1, at 517-22 (discussing whether consensus in ratings of seriousness is attributable to shared conceptions of the conduct itself or shared perceptions of the legal severity of acts).

[7] *See* Rossi & Henry, *supra* note 1, at 493-94 (discussing an aggregation bias and the consequence of using mean ratings of groups to measure consensus).

[8] For reviews of the finding of consensus, see Miethe, *supra* note 1, at 515-17; Turner, *supra* note 4.

[9] This paper focuses on measuring consensus between groups on all acts or subsets of acts. The appropriate statistical technique for measuring the amount of consensus on a particular act is a measure of dispersion such as a standard deviation, coefficient of relative variation, or interquartile ranges. For a discussion of these latter measures, see Chilton & DeAmicis, *Overcriminalization and the Measurement of Consensus,* 59 SOC. & SOC. RES. 318 (1975); Miethe, *supra* note 1, at 521-22.

TABLE 1

CLASSIFICATION OF TYPES OF CONSENSUS AND
APPROPRIATE DATA-ANALYTIC STRATEGIES FOR
EACH TYPE OF CONSENSUS*

| | | TYPE OF RATINGS | |
		Relative	Absolute
DEGREE OF ITEM INCLUSION	Global	Correlation: Pearson's r Spearman's rho	Regression(a=0 b=1) t-tests and Correlation Hotelling's t^2
	Local	Correlation: Pearson's r Spearman's rho	Regression(a=0 b=1) t-tests and Correlation Hotelling's t^2

Notes: *See text for description of types of ratings and degree of item inclusion. The data-analytic strategies are generally appropriate for either individual or aggregate levels of analysis. With aggregate measures it is also necessary to examine the magnitude of within-item variability for each group.

The first dimension of the typology (type of ratings) differentiates between statements about the relative ordering of items in a scale and the actual magnitude of the scores assigned by different social groups. Although this distinction often becomes muddled, past studies typically use *relative consensus*,[10] which is assessed by correlating the mean ratings of different groups. A correlation greater than .70 has sometimes been used as an indicator of this type of consensus.[11] The effect of aggregation, which inflates the magnitude of the association, is gauged by comparing the correlation between group means with either the average individual-individual correlation[12] or the average individual-group correlation.[13] The results of such comparisons tend to support the conclusion that, even at the individual level, there is remarkable agreement in the relative ratings of seriousness.

Absolute consensus is far more restrictive than relative consensus. It

---

[10] *See* T. SELLIN & M. WOLFGANG, THE MEASUREMENT OF DELINQUENCY (1964); Cullen, Link & Polanzi, *The Seriousness of Crime Revisited: Have Attitudes Toward White-Collar Crime Changed?*, 20 CRIMINOLOGY 83 (1982); Figlio, *The Seriousness of Offenses: An Evaluation of Offenders and Nonoffenders*, 66 J. CRIM. L. & CRIMINOLOGY 189 (1975); Rossi, Waite, Bose & Berk, *supra* note 2; Sinden, *Offenders' Gender and Perceptions of Crime Seriousness*, 1 SOC. SPECTRUM 39 (1981).

[11] Rossi, Waite, Bose & Berk, *supra* note 2, at 227.

[12] *See* McClearly, O'Neil, Epperlein, Jones & Gray, *Effects of Legal Education and Work Experience on Perceptions of Crime Seriousness*, 28 SOC. PROBS. 276 (1981).

[13] Rossi, Waite, Bose & Berk, *supra* note 2, at 235-36.

refers to cases in which the ordering of items and the numerical scores assigned to each item are similar across and within social groups. Two different analytic strategies are available to measure this type of consensus.[14] First, a regression analysis can be performed on the mean scores of different groups, with absolute consensus being supported if the intercept and slope of the regression equation are not significantly different from the hypothesized values (i.e., a=0; b=1). A second approach involves t-test comparisons of group means for each item and the correlation of mean ratings across all items. If there are no significant differences in group means *and* the mean scores are highly correlated, then the results support absolute consensus. With both strategies for evaluating absolute consensus, however, it is necessary to examine the degree of variability around the mean ratings as evidence of within-group dissensus.

It is important to note that the existence of absolute consensus, as defined here, also reflects relative consensus. In contrast, one might achieve perfect relative consensus with little or no absolute agreement. For example, this might occur if male and female scores are perfectly correlated, but females considered all acts as more serious than males.

The second dimension (degree of item inclusion) is what Rossi and Henry[15] refer to as a distinction between "global" and "local" consensus. *Global consensus* is defined as agreement in ratings across the total list of items included in a scale. *Local consensus*, on the other hand, refers to agreement on items that are either close together on the scale (e.g., the most serious crimes) or of a particular type (e.g., white-collar violations or violent crimes). Although a few exceptions exist,[16] previous research has focused on global consensus.

Table 1, combining these two dimensions, reveals four distinct types of consensus: global relative, global absolute, local relative, local absolute. Past studies reporting consensus in ratings of seriousness usually mean the *global relative* type. In particular, these studies usually evaluate consensus by computing correlation coefficients on the basis of

---

[14] There are, of course, other strategies for measuring absolute consensus. One useful measure is the T statistic for testing the differences between mean vectors. *See* D. MORRISON, MULTIVARIATE STATISTICAL METHODS (1976). This test statistic evaluates the likelihood that two sets of sample means were generated from the same population. A number of test statistics (e.g., standardized residuals) are also available in SPSS-X and SAS to detect influential points and outliers in regression models. These latter statistics are useful to locate the source of deviations from, and contributions to, absolute consensus. A strategy for measuring absolute consensus on individual acts is to examine the percent of variation in individual ratings of the seriousness of a particular act explained by respondents' demographic profile. For an application of this approach to measure consensus between groups, see Hawkins, *Causal Attribution and Punishment for Crime*, 2 DEVIANT BEHAV. 207 (1981).

[15] Rossi & Henry, *supra* note 1, at 494.

[16] *See* Figlio, *supra* note 10; Hawkins, *supra* note 14; Rossi & Henry, *supra* note 1.

mean ratings across the entire range of items. Using this strategy, it is common to obtain high correlations between group means (r's > .90). There are, however, two major problems with this strategy for measuring consensus. First, an aggregate-level correlational approach ignores individual variability. Second, agreement on the ratings of only a few items, especially those at the extremes of the scale that provide the largest contribution to the covariance, may mask widespread disagreement on the ratings of the majority of items.[17] This latter problem exists regardless of the unit of analysis. A preoccupation with this type of consensus, to the neglect of other types and measures, might severely distort the extent of agreement among social groups in their ratings of seriousness of crimes.

III. METHODS

Data collected by Rossi et al.[18] was reanalyzed to demonstrate the various measures of consensus at the individual and aggregate levels of analysis. The Rossi study consists of a block quota sample of 125 white and 75 black Baltimore residents. The interviewees rated subsets of 140 criminal offenses on a nine-point scale ranging from most (9) to least (1) serious. Two lists of eighty acts were used, with twenty acts common to both lists and sixty acts being different on each list.[19] There were sixty-three whites and thirty-seven blacks assigned to rate acts in the first list; sixty-two whites and thirty-eight blacks rated the other acts.

The criminal acts were derived by transforming the *Uniform Crime Reports'*[20] listing of general crime categories into specific acts.[21] To assess the extent of local consensus, the acts were regrouped into the following categories: personal, violent crimes;[22] property offenses;[23] white collar offenses;[24] victimless crimes;[25] and public order violations.[26] Mul-

---

17 For related discussions, see Chilton & DeAmicis, *supra* note 9; Miethe, *supra* note 1, at 517-18; Rossi & Henry, *supra* note 1, at 493-95. As an example of how a preoccupation with global relative consensus can mask dissensus among groups, Figlio found that the geometric mean ratings of institutionalized samples of adult and juvenile offenders were highly correlated (r=.962). *See* Figlio, *supra* note 10, at 194. Yet, when the assaults were excluded, this correlation dropped to .59 (these correlations were calculated from Table IV).

18 Rossi, Waite, Bose & Berk, *supra* note 2.

19 For a full discussion of the sample and methodology, see Rossi, Waite, Bose & Berk, *supra* note 2, at 225-27.

20 FEDERAL BUREAU OF INVESTIGATION, UNIFORM CRIME REPORTS FOR THE UNITED STATES AND ITS POSSESSIONS 5-155 (1982).

21 *See* Rossi, Waite, Bose & Berk, *supra* note 2, at 228, Table 1.

22 Thirty personal, violent crimes, such as murder, assault, and rape, were included in this category.

23 Twenty-six property offenses, such as breaking/entering, shoplifting, and burglary, were included in this category.

24 Thirteen white collar offenses, such as embezzlement, stock fraud, and repair and service overcharge, were included in this category.

TABLE 2

CORRELATION AND REGRESSION ANALYSES OF MEAN
RATINGS OF WHITES AND BLACKS FOR ALL ACTS
AND SUBSETS OF ACTS*

Acts	N of (acts)	r	Regression Model**	F-ratios for Ho: a=0 and b=1***
ALL ACTS	(140)	.897	$Y_w = -.369 + .994X_b$ (.052) (.042)	$F_{2,138} = 31.11$****
VIOLENT	(30)	.936	$Y_w = -.558 + 1.053X_b$ (.063) (.075)	$F_{2,28} = 3.55$
PROPERTY	(26)	.667	$Y_w = -.007 + .897X_b$ (.138) (.205)	$F_{2,24} = 11.39$****
WHITE COLLAR	(13)	.932	$Y_w = -.766 + 1.096X_b$ (.094) (.091)	$F_{2,11} = 2.55$
VICTIMLESS	(12)	.933	$Y_w = -1.19 + 1.049X_b$ (.158) (.130)	$F_{2,10} = 14.62$****
PUBLIC ORDER	(29)	.837	$Y_w = -.124 + .973X_b$ (.145) (.122)	$F_{2,27} = 1.73$

Notes: *The number of respondents used in the computation of each mean rating
ranged from 37 to 75 for blacks and 62 to 125 for whites.

**In each regression model, Y_w refers to the predicted mean ratings of whites and
X_b refers to the mean ratings of blacks. The standard errors for the t-tests of
the separate hypotheses that a=0 and b=1 are in parentheses ( ).

***The F-tests are based on the Formula: $$\frac{\Sigma SS_1 - \Sigma SS_2/2}{\Sigma SS_1/N-2}$$

where ΣSS_1 refers to the sum of squared errors for the estimated model in
which the hypotheses are imposed and ΣSS_2 refers to the sum of the squared
errors for the estimated model. For a discussion of this F-test, see Namboodiri,
Carter & Blalock, *supra* note 29, at 164-66.

****Significant at p < .01, i.e., reject H_0 of absolute consensus.

tiple expert judges were used for the subclassification of acts.[27] Thirty
acts that received multiple classifications were excluded from the analy-
sis of the extent of local consensus. For purposes of illustrating the differ-
ent kinds of consensus, correlation, t-tests, and regression analyses were
restricted to comparisons of the ratings of black and white
respondents.[28]

---

25 Twelve victimless crimes, such as homosexuality, prostitution, and selling or using her-
oin, LSD, or pep pills, were included in this category.

26 Twenty-nine public order violations, such as income tax violations, loitering, and join-
ing a riot/demonstration, were included in this category.

27 The multiple expert judges were four individuals employed in the field of criminal
justice. These judges independently classified each act into one of the five categories.

28 Because the major purpose of this study is to illustrate how to assess different kinds of

IV. RESULTS

MEASURES AND TYPES OF CONSENSUS

Correlation and regression analyses of the mean ratings of blacks and whites were performed on the total set of acts and particular subsets of acts. Table 2 summarizes the extent of the four types of consensus.

The typical measure of consensus (global relative) requires correlating group means across all acts. As shown in Table 2, the mean ratings of the races across all acts exhibit a high degree of global relative consensus (r=.897). Black and white ratings of violent, white-collar, and victimless crimes are also highly correlated (r's > .930), reflecting local relative consensus on these subsets of acts. The mean ratings for public order crimes are moderately related (r=.837), but there is relatively little local relative consensus on the ratings of property offenses (r=.667).

If the mean ratings of blacks and whites exhibit absolute consensus, the intercept and unstandardized regression coefficient for the model $Y_w = a + bX_b$ should not depart significantly from their hypothesized values (a=0 and b=1). The hypothesized model (y = 0 + 1X = X) was compared with the estimated model (y = a + bX) and the reduction of errors in prediction was assessed.[29] If the reduction in prediction errors is insignificant, the hypothesis of absolute consensus between mean ratings of blacks and whites would not be rejected. Table 2 displays the regression equations and the tests of absolute consensus for all acts and subsets of acts.

As shown in Table 2, the data does not support the hypothesis of global absolute consensus ($F_{2,138}=31.11$, p < .01). Although the slope is not significantly different from 1, the intercept reveals that, overall, blacks considered the acts more serious than whites. For violent, white-collar, and public order offenses, however, the reduction of prediction errors is insignificant, indicating local absolute consensus for these acts.[30] In contrast, the data does not support the hypothesis of local

---

consensus, racial differences were examined because the use of quota sampling in the original study guaranteed a sufficient number of blacks and whites in the sample. In terms of all possible bivariate comparisons of demographic characteristics, however, it was also expected that such groups should exhibit the most dissensus based on previous research and theories of racial differences in criminal and non-criminal conduct. For an investigation of racial differences in ratings of severity of punishments, see Hawkins, *supra* note 14.

[29] This test of absolute consensus is identical to a test of reliability. An F-test is used to assess the reduction in prediction errors in the two models because two constraints are imposed on the hypothesized model (i.e., a=0 and b=1). For a discussion of this test and the computation of the sum of squared errors in prediction for the models, see K. NAMBOODIRI, L. CARTER & H. BLALOCK, APPLIED MULTIVARIATE ANALYSIS AND EXPERIMENTAL DESIGN 164-66 (1975).

[30] The results of the regression analyses are similar to those obtained by using both t-tests for differences in means and correlations between the mean ratings of blacks and whites. No

TABLE 3

THE PREDICTIVE POWER OF RATER'S DEMOGRAPHIC
PROFILE ON RATINGS OF THE SERIOUSNESS OF
INDIVIDUAL ACTS*

(N) of Acts**	R^{2***}	SD of R^2's	Range of R^2's	%R^2's > .10	Most Common Predictors****
VIOLENT (30)	.088	.039	.013—.157	26.7% (8/30)	Age (4) Race × Ed (3) Race × Sex (2)
PROPERTY (26)	.144	.044	.056—.252	88.5% (23/26)	Ed (4) Race (3)
WHITE COLLAR (13)	.108	.049	.044—.231	53.9% (7/13)	Age (3) Race × Sex (2) Race × Ed (1)
VICTIMLESS (12)	.187	.086	.094—.352	91.7% (11/12)	Race (2) Age (2)
PUBLIC ORDER (29)	.137	.057	.047—.308	75.9% (22/29)	Race × Sex (9) Race × Ed (4) Sex × Ed (4) Race (2) Ed (3)

Notes: *The demographic characteristics of the respondent included in this analysis
were race, sex, education level, age, victimization experience, and all two-way
interactions between race, sex, and education. See text for a coding of these
variables.

**N of acts refers to the number of acts included in each subset.

***R^2 refers to the average proportion of variation in individual ratings of each act
explained by respondent characteristics. It was computed by summing up the
R^2's for each act and dividing by the number of acts in its respective subset.

****The most common predictors are those demographic variables that were found
to be significant in the analysis of each act's seriousness. The number of acts
within each subset that contained these variables as significant predictors are in
parentheses ( ).

absolute consensus on ratings of property or victimless crimes.

A series of regression analyses were performed to assess the level of
absolute consensus on individual ratings of each act. In particular, the
proportion of variation in each act's seriousness accounted for by the
respondent's demographic profile was examined. The demographic
variables used in this analysis were race (black, white), sex, education

---

significant differences in means are observed for the white-collar offenses, and only one of the
thirty t-tests are significant for the violent acts. The mean ratings for these acts are also
highly correlated (see Table 1), reflecting local absolute consensus for these acts. On the other
hand, although the mean ratings are highly correlated, 50% (6/12) of the t-tests for the vic-
timless crimes are significant, suggesting local relative but no absolute consensus in black and
white ratings of these acts.

TABLE 4

INDIVIDUAL-LEVEL ANALYSIS OF RELATIVE CONSENSUS FOR ALL 140 CRIMES AND SUBSETS OF CRIMES FOR THE TOTAL SAMPLE AND EACH RACE

Acts ($\bar{x}$ of acts)*	N**	$\bar{r}_{i\bar{x}}$***	sd of r's	range of r's
140 ACTS (6.27)				
Total	193	.567	.189	−.05 - .87
Whites	120	.549	.174	.01 - .81
Blacks	73	.482	.188	.00 - .79
30 VIOLENT (7.39)				
Total	192	.535	.283	−.33 - .93
Whites	120	.536	.274	−.30 - .96
Blacks	72	.478	.288	−.27 - .91
26 PROPERTY (5.87)				
Total	193	.419	.304	−.50 - .91
Whites	120	.355	.268	−.49 - .78
Blacks	73	.215	.340	−.58 - .83
13 WHITE COLLAR (5.77)				
Total	185	.452	.434	−.99 - .99
Whites	115	.424	.424	−.99 - .98
Blacks	70	.371	.443	−.88 - .98
12 VICTIMLESS (6.17)				
Total	185	.574	.329	−.63 - 1.00
Whites	117	.544	.360	−.63 - 1.00
Blacks	68	.533	.314	−.64 - .95
29 PUBLIC ORDER (5.13)				
Total	193	.545	.270	−.40 - .94
Whites	120	.538	.242	−.35 - .94
Blacks	73	.419	.285	−.31 - .93

Notes: *$\bar{x}$ of acts refers to the grand mean of all acts in that subset for all respondents.

**N refers to the number of individuals on which the correlation is based. The sample sizes vary because of the exclusion of individuals in each subset who had no variation in their ratings of acts.

***$\bar{r}_{i\bar{x}}$ refers to the average correlation between each individual's ratings and the mean ratings of their respective group. For the total row, $\bar{x}$ is the mean for all individuals, for whites it is the mean for all whites, and for blacks it is the mean rating given by all blacks.

(high school, college), age, victimization experiences (yes, no), and a set of interaction terms for race and sex, race and education, and education and sex. Table 3 summarizes the most common predictors, the average R^2, and other descriptive statistics for each subset of acts.

As shown in Table 3, there is a great deal of variation in the predictive power of demographic variables within and between each subset of acts. Relatively little of the variation in ratings of violent, white-collar,

or public order offenses is explained by the respondent's demographic profile,[31] a finding consistent with the previous analyses of racial differences. For property and victimless crimes, on the other hand, a sizeable proportion of the variation in each act is accounted for by the demographic characteristics of the respondent, suggesting little absolute consensus in ratings of these acts. The most important predictors for these acts are the race, educational level, and age of the respondent: older people considered victimless crimes more serious than did younger people; those with some college education considered property crime less serious than did those with no college education; and whites considered property crime more serious and victimless crimes less serious than did blacks. Thus, although characteristics of the offense have been shown to influence ratings of seriousness,[32] characteristics of the respondent also have a noteworthy impact on the ratings of victimless and property crimes, explaining at least 10% of the variation for at least 80% of these acts.

MEASURES OF INDIVIDUAL LEVEL CONSENSUS

The effect of aggregation on the findings of relative consensus was evaluated by computing the average correlation between each individual's scores and the mean ratings for all individuals. These average individual-mean correlations were also computed separately within each racial grouping. Table 4 presents the results of these individual-level comparisons.

The average individual-mean correlation is .567, whereas separate analysis by race reveals greater agreement among whites than blacks (r_{ix} = .549 and .482, respectively). Thus, the extent of global relative consensus is reduced from .897 to .567 as the unit of analysis shifts from groups to individuals. Although aggregation effects are present, a modest degree of this type of consensus remains at the individual level.[33]

The possibility that global relative consensus is attributable to agreement at the extremes of the scale (i.e., the most and least serious acts) was examined by performing separate analyses for subsets of

---

[31] Significant differences by several categories exist even for some acts in these categories, however.

[32] *See* T. SELLIN & M. WOLFGANG, *supra* note 10; Rossi, Waite, Bose & Berk, *supra* note 2; T. Miethe, The Social and Methodological Determinants of Consensus in Public Evaluations of Crime (unpublished manuscript presented at the Annual Meetings of the American Society of Criminology, Denver, Colorado, Nov. 10-12, 1983).

[33] When the average individual-individual correlation r_{ij} is used as the measure of relative consensus at the individual level, the ordering of the subsets of acts is similar to that reported in Table 4, but the magnitude of the consensus is far less. The corresponding values for the r_{ij}'s are: .311, .279, .254, .251, .339, and .292. Support for relative consensus at the individual level is less convincing when these latter coefficients are used.

crimes. As shown in Table 4, the most serious acts (violent and victimless crimes) and least serious acts (public order crimes) exhibit the highest levels of individual agreement and make the largest contribution to the overall level of global relative consensus. In contrast, acts considered moderately serious (property and white-collar crimes) elicit the least individual agreement.[34] Comparisons of correlations and standard deviations within each race reveal greater relative consensus among whites than blacks for all acts and subsets of acts. Thus, it appears that a preoccupation with global relative consensus can mask systematic variation between and within social groups in their ratings of subsets of acts even at the individual level of analysis.

V. Discussion and Implications

The results of this study illustrate that different analytic strategies yield different conclusions about the extent of consensus in blacks' and whites' ratings of crime seriousness. Using a correlational approach, the mean ratings of the races exhibit a high level of relative consensus across all acts and for the violent, white-collar, victimless, and public order subsets of acts. There is a moderate amount of global relative consensus even at the individual level, but this is due largely to agreement on the ratings of the most serious acts (violent, victimless) and least serious acts (public order). Regression analyses reveal absolute consensus on the ratings of violent, white-collar, and public order offenses, but not on the victimless crimes. Neither absolute nor relative consensus is observed in the ratings of property offenses. Similar results are found when other demographic characteristics of the respondent are considered. Individual-level comparisons and the analysis of within-group variation reveal greater agreement among whites than blacks.

Past researchers assert that studies of public evaluations of crime are relevant to a number of policy issues. For example, particular states are requesting these studies as a basis for equating legal proscriptions with public opinion,[35] and as a means of allocating prosecutorial resources on criminal cases.[36] Implicit in these applications is the belief that public consensus in ratings of criminal acts exists. Unfortunately, it is unclear how much and what type of agreement qualifies as public consensus. Given the existence of absolute consensus in ratings of vio-

---

[34] The grand means for all acts in each subset for all respondents are presented in parentheses in Table 4. The relatively high grand mean for the victimless crimes (6.17) is largely due to the inclusion of selling heroin and LSD as victimless crimes. Excluding these acts reduces the average individual-mean correlation from .574 to .525 for the victimless crimes.

[35] Rossi & Henry, *supra* note 1, at 495-502.

[36] Roth, *Prosecutor Perceptions of Crime Seriousness*, 69 J. CRIM. L. & CRIMINOLOGY 175 (1978).

lent, white-collar, and public order offenses, these findings might serve as a rational basis for adjusting statutory penalties for these acts.[37] It is still uncertain, however, how seriousness ratings can be calibrated into criminal sanctions. Further, as illustrated in Table 3, racial and other differences exist even for these crimes. Thus, adjusting statutory penalties to be congruent with ratings of seriousness is questionable because these penalties would not accurately represent public sentiments even under conditions of absolute consensus.

The utility of studies of crime seriousness for policymakers appears to depend upon the identification of factors that influence ratings of crime. In other words, the finding of absolute consensus in ratings of subsets of acts becomes meaningful to judicial or legislative decision-making only when factors influencing these ratings are identified and tested.

A review of past studies suggests that respondents consider a variety of circumstances and characteristics of the offense when rating the seriousness of crimes. For example, in a previous analysis of the Baltimore data, Miethe[38] found that over 71% of the variation in mean seriousness ratings was explained by the type of harm, intentionality of the act, and whether victim consent was implied in the offense descriptions. The degree of harm (e.g., amount of physical injury or monetary loss) has also been shown to influence the seriousness of acts.[39] Further, Sykes and West[40] found that respondents consider the amount of harm suffered, the degree of intent, motive, and characteristics of the victim and the offender when evaluating an act. Unfortunately, none of the previous analyses of the determinants of seriousness have considered the unique and interactive effects of these variables. The use of factorial designs, in which combinations of these factors are introduced in the offense descriptions, should be encouraged as a strategy for examining the unique and interactive effects of these variables on ratings of seriousness.

An examination of the ratings given to particular acts within each subset of crimes reveals some other factors that might explain ratings of seriousness. For example, the high level of relative consensus on aggregate and individual ratings of victimless crimes is largely due to agreement on the most serious acts (e.g., selling/using heroin or selling LSD), acts that are typically viewed as harmful to users and others. Similarly,

---

[37] The existence of absolute consensus means that relative consensus also exists. *See* text accompanying Table 1.

[38] T. Miethe, *supra* note 32.

[39] *See* Riedel, *Perceived Circumstances, Inferences of Intent, and Judgments of Offense Seriousness*, 66 J. CRIM. L. & CRIMINOLOGY 201 (1975); T. SELLIN & M. WOLFGANG, *supra* note 10.

[40] G. Sykes & S. West, The Seriousness of Crime: A Study of Popular Morality (unpublished manuscript presented at the Annual Meetings of the Eastern Sociological Society, 1978).

even though absolute consensus exists in ratings of violent and white-collar offenses, a closer examination of particular acts suggests some additional factors influencing judgments of seriousness. In particular, blacks considered twenty-two of the thirty violent offenses as slightly more serious than whites. Yet, all impulsive slayings and most of the violent offenses against strangers were considered slightly more serious by whites, suggesting that particular characteristics of the offense and the participants are differentially evaluated by blacks and whites. White-collar crimes involving victimization of consumers (e.g., using inaccurate scales to weigh meat, fixing prices on products, false advertising, and refusal to make rental repairs) were considered slightly more serious by blacks. In contrast, whites considered crimes against business (e.g., embezzlement and illegal interest rates) as being slightly more serious than did blacks, suggesting that the relevance of the act to one's personal experiences and the likelihood of exposure to it influence ratings of seriousness.[41] Thus, the application of seriousness studies to public policy decisions appears to be premature until factors influencing ratings of seriousness and explanations for consensus have been identified and tested.

The distinctions between types of consensus illustrated in this paper are directly relevant to policy and theoretical issues. For example, a finding of relative consensus in ratings of "powerless" and "elite" groupings is largely irrelevant as a basis for evaluating consensual or conflict models of law creation.[42] If a conflict perspective is supported, however, one would expect differences in the magnitude of ratings assigned by social groups to particular acts (e.g., victimless crimes, white-collar violations), suggesting that the presence or absence of local absolute consensus is the type of consensus that is the most appropriate for addressing these issues.[43] Similarly, attempts to translate seriousness ratings into legal sanctions would best be accomplished only under conditions of absolute consensus. In contrast, when using these studies as a basis for allocating police or prosecutorial resources to fight crime or for cross-cultural comparisons, a finding of global or local relative consensus is probably sufficient.[44]

---

41 For a related discussion, see Hawkins, *supra* note 14.

42 The consensus model of law creation posits that law reflects public sentiments about particular conduct, whereas the conflict model proposes that law is a reflection of the morality of those few who are in a position to have their interests codified as law. A finding of relative consensus is largely irrelevant to an evaluation of these models because it is the *magnitude* of ratings given particular acts, rather than their relative ordering, that would illustrate a disparity between the powerless and powerful.

43 For a test of these models, see Sinden, *Perceptions of Crime in Capitalist America: The Question of Consciousness Manipulation*, 13 Soc. Focus 75 (1980).

44 Again, the use of these studies for addressing theoretical or programmatic issues is pre-

The distinction between relative and absolute agreement also should be useful for evaluating theories of criminal behavior. For example, a variety of subcultural theories posit that different rates of criminal behavior among social groups is the result of a lifestyle that is not necessarily opposed to the mainstream ideology but, rather, places greater importance on particular values. In other words, different social groups may agree on the relative importance of attitudes, values, or behavior, but it is their absolute importance that contributes to deviant conduct.

The subculture of violence thesis,[45] for example, posits that within particular social groups there is a value system that encourages a violent response to threatening stimuli. It argues further that there is a differential perception of anger-provoking stimuli, with "trivial" altercations becoming fighting words among particular groups. In terms of the relative and absolute distinction, one would expect blacks and whites to agree in their relative ratings of violent acts, but lower-class, black males consider these acts more serious than other social groups.[46] Similarly, the notion of "focal concerns" of lower class youth[47] and the image of "lower class value stretch"[48] are suggestive of a distinction between agreement among social groups in the relative ordering of values and disagreement among those groups in their absolute endorsement of them.

Given the volume of studies on public images of crime, it is surprising that a definitive statement about the extent of agreement in public ratings of the seriousness of crime cannot be made. The fact that various degrees and types of consensus can be found in the same data, however, is not necessarily bothersome. It simply requires researchers to pay more attention to the variety of techniques available to measure consensus *and* to specify the type of consensus that is most appropriate for the particular research question. Although a finding of absolute consensus appears to be most relevant for policy issues and theoretical development, relative consensus might be adequate for some research purposes.

---

mature until factors influencing ratings of seriousness have been clearly identified and tested. Once this is done, however, the distinction between types of consensus seems to be relevant for addressing these issues. The relevance of crime seriousness studies as a basis for evaluating the competing theories of law has been questioned. *See* Rossi & Henry, *supra* note 1.

[45] M. WOLFGANG & F. FERRACUTI, THE SUBCULTURE OF VIOLENCE (1967).

[46] Although an adequate measure of social class was absent in the original study, the finds reported earlier are not consistent with the subculture of violence thesis. In particular, the findings of absolute consensus between blacks and whites in their ratings of the violent offenses and that white, college-educated respondents considered these acts *more* serious than less-educated blacks are inconsistent with the basic assumption underlying the subculture of violence argument.

[47] Miller, *Lower-Class Culture as a Generating Milieu of Gang Delinquency*, 14 J. SOC. ISSUES 5 (1958).

[48] Rodman, *The Lower Class Value Stretch*, 42 SOC. FORCES 205 (1963).

This research should be useful for future studies of public evaluations of crime. If researchers adopt multiple strategies for measuring different types of consensus, we will be more confident about the nature and extent of public agreement on ratings of crime. By investigating different types of consensus and factors that contribute to these ratings, researchers will be in a better position to determine whether studies of crime seriousness are useful for judicial decisionmaking and for evaluating different theories of law and criminal behavior.