
ON MEASURING A NORM: SHOULD THE
PUNISHMENT FIT THE CRIME?

by V. Lee Hamilton and Steve Rytina

University of Michigan

May, 1978

CRSO Working Paper #175

Copies available through:

The Center for Research on
Social Organization
The University of Michigan
Perry Building
330 Packard
Ann Arbor, Michigan 48109

On Measuring a Norm:
Should the Punishment Fit the Crime?*

by

V. Lee Hamilton
Steve Rytina

University of Michigan

*The research and analysis reported here were supported by N.S.F. grant #SOC 76-01113 to Hamilton and by the Center for Studies in Criminology and Criminal Law, University of Pennsylvania. We wish to thank the Laboratory for Behavioral Research, Political Science Department, S.U.N.Y. at Stony Brook, and particularly David Cross and Laurence Rotkin, for making the research possible. We also thank Marvin Wolfgang for his support and encouragement, and William Form and Robert Zajonc for their helpful comments on a previous draft.

Order of authorship is alphabetical, as each author contributed equally to the final product. Requests for reprints should be sent to V. Lee Hamilton, Sociology Department, University of Michigan, Ann Arbor, MI 48109.

ABSTRACT

Normative consensus-- or its absence-- has long been a central concern in sociology. The present paper explores the possible meanings of consensus and tests the degree of cognitive consensus on the norm of just deserts (i.e., "letting the punishment fit the crime"). It is argued, first, that consensus on complex norms of exchange, justice, and sanctioning should be tested using a combination of within-respondent and between-respondent techniques. Such tests can (1) simultaneously reveal the presence of consensus on the principle involved and on the specific social stimuli presented; (2) facilitate demographic comparisons on adhering to principle or agreeing on facts; and (3) reveal conflicts between these two versions of consensus. For testing the norm of just deserts, ratio scale measures of crime seriousness and punishment severity were employed, and a formula derived from both equity theory and psychophysics was utilized in model-fitting. Results from a sample survey indicated dramatically strong use of the principle of just deserts by members of the public, but less adherence to just deserts by demographically disadvantaged (low income or black) respondents. A path model of the relation between aggregate and individual scores demonstrated a fundamental tension between the two versions of normative consensus, in that the more respondents used the principle the more systematically they deviated from the group average response. The paper concludes by discussing the implications of both the substantive findings and the new methodologies employed for the understanding of normative consensus and the assessment of justice norms.

ON MEASURING A NORM: SHOULD THE PUNISHMENT FIT THE CRIME?

"Norm" is a throwaway word in modern sociology. Norms tend to be invoked rather than measured; the statement that something is a norm is more often the conclusion of an investigation than its starting point. Nevertheless, there is general agreement that a norm is a "rule, standard, or pattern for action" (Williams, 1968, p. 204). The notion that such rules exist is pervasive in sociology. Yet the frequency with which they are invoked is not matched by clarity of definition while doing so. Within the general definition noted above, two nuances of meaning are typically left undistinguished: norms as social structural entities and as cognitive structures. As social structural entities, norms can be identified by patterns of behavior. Nadel (1957) distinguished three properties of norms in this sense: statistical frequency, mechanisms for sanctioning, and codification.¹ But the rules that regulate behavior are also presumably to be found in the heads of the people who follow them. Individual members of a society can both give statements of what they see norms to be and evidently apply their images of norms in their personal evaluations of specific observed behaviors. These individual adherences represent norm in its second sense: as a cognitive structure in the heads of the society's participants. This second notion of the meaning of norm, as a cognitive principle for ordering one's

perceptions and evaluations, is the object of the present study.

For norms as cognitive structures, a central empirical question is the existence and degree of consensus. The simplest version of a cognitive norm involves an isolated rule associating an act with an evaluation or sanction. Isolated statements of the relation between act and evaluation can be the objects of agreement or disagreement by members of society. A ready measure of consensus among members would be similarity of their responses on a sample survey. The notion of consensus is a good deal less clear for cognitive principles at a higher level of abstraction, those which serve to unite and order disparate objects of moral evaluation. In general, norms governing exchange, resource allocation, and sanctioning itself-- norms of justice and fairness-- are such higher level abstractions. As an example, the norm of reciprocity (Gouldner, 1960) is a higher level cognitive principle which connects disparate acts of altruism in a relation of moral entailment or obligation. Consensus on such norms is considerably more difficult to evaluate.

Why is consensus so complex for norms of justice? As a principle, each entails the assessment of a category of inputs, a category of outputs, and the relation between them. For example, the principle "equal pay for equal work" involves evaluations of work, pay and "equal." To be assessed as a cognitive structure, such a norm should be tested first at the level of the individual respondent. Such a test would involve assessments by the respondent of the relative value of inputs, the relative value of outcomes, and the appropriate match between the two. However, individual respondents could all agree at the level of the principle, all matching what they take to be

equal input with output, while totally disagreeing among themselves on any specific instance of the work-pay relation. Such disagreement would follow from lack of group consensus about the inputs, the outputs, or the exact transformation rule necessary to achieve equal matching. Thus societal consensus at the level of principle can co-exist with societal dissensus at the level of application. In summary, the investigation of an extremely important category of social norms, those involving higher level abstractions like justice, entails a complex evaluation of the presence of consensus and evokes potentially different versions of consensus itself.

The goal of the present study is the testing of consensus on a single norm, that of the fair sanctioning of criminal acts. Based on the prior literature, we take the currently dominant ideology of fair sanctioning to be the philosophical principle of just deserts-- more musically expressed by Gilbert and Sullivan as the notion that one should "let the punishment fit the crime." We investigate the degree of individual adherence to the principle; the extent to which individual adherence is accompanied by aggregate consensus between individuals; and the extent to which individual adherence is explained by aggregate consensus. First, however, the investigation is empirically grounded in prior social science investigations of consensus on criminal justice. Second, it is theoretically and mathematically grounded in the most relevant developed approach to perceptions of justice, social psychological equity theory (cf. Berkowitz & Walster, 1976; Walster, Walster, & Berscheid, 1973).

Prior Studies and Codification of Fitting the Punishment to the Crime

The notion that the punishment should fit the crime has served as the unexamined backdrop for the most widely studied issue in consensus on criminal justice: measurement of the seriousness of crime. In the landmark study of Sellin and Wolfgang (1964), strong agreement among judges, police and students was observed on psychophysical measures of crime seriousness. Rossi, Waite, Hose, and Berk (1974) extended the examination of consensus to a quota sample of Baltimore, using different scaling methods but arriving at quite similar results. Numerous other studies have shown strong evidence both within this culture and cross-culturally of a relatively invariant ordering of crimes on relative seriousness.²

But is crime seriousness itself thereby a norm? Rossi et al. (1974, p. 237) certainly treat it as such, referring in their conclusions to "norms defining how serious various criminal acts are considered to be..." "general normative trends," and "the normative structure." Yet crime seriousness is only normative - either cognitively or social structurally - if it then implies something about action. What Rossi et al. implicitly include is the untested principle that the punishment should fit the crime. They make this more explicit in introducing their study, by saying that societal consensus "should be reflected in the criminal code, the behavior of judges and juries, and the actions of law enforcement agencies" (p.224) -- in other words, that judges, etc. respond differentially to crimes according to their seriousness. Unless they wish to argue that more serious crimes are felt to deserve less punishment, Rossi et al. must mean what Gilbert and Sullivan did -- that more serious crimes should be and are responded to with more punishment.

Sellin and Wolfgang (1964) also implicitly assumed the principle of just deserts in validating their crime seriousness scale. In one of their four modes of validation, they correlated the average seriousness of crimes according to their respondents' ratings with the maximum punishment then allowed in the Pennsylvania penal code, arriving at a correlation of .93 between their aggregate seriousness and the Pennsylvania maximum punishment. By treating this as a scale validation, rather than as evidence that crime seriousness was merely an alternative measure of punishment severity, Sellin and Wolfgang assumed something: that crime seriousness was the input, punishment severity the output, and that the norm was one of fitting the punishment to the crime.³

It is not necessarily the case that the punishment should fit the crime. Among the many competing approaches to why criminal behavior should be punished, the major alternative view is the rehabilitationist approach. This approach argues that the punishment should be forward-looking, fitting the criminal rather than the offense, and geared to the modification of the offender's future behavior. This approach is not necessarily more lenient, being associated with such practices as indeterminate sentencing, but it is an alternative. Conversely, the principle that the punishment fit the crime should not be identified solely with the Biblical retributivist "an eye for an eye..." Instead, some notion of fit between seriousness of offense and severity of punishment can be seen as consistent with deterrence and incapacitation approaches as well as a retributive approach to punishment (von Hirsch, 1976). In general, it has recently been argued that the philosophical notion of just deserts, according to which punishment should be graded according to the seriousness of the offense, is the fundamental

principle of fairness in punishment (Hart, 1968; von Hirsch, 1976). It is incorporated into all approaches to why individuals should be punished except for rehabilitation; it is also observable in virtually all codifications of sentencing practices. In short, desert would appear to be the dominant ideology of punishment in Anglo-American law and society, and thus is likely to be the dominant consensually held version of criminal justice among members of the public.

In the cognitive representation held by members of the public, just deserts is simply the Gilbert and Sullivan notion that the punishment should fit the crime, so far assumed rather than tested in investigations of social consensus regarding criminal justice. Testing the norm of just deserts as a cognitive structure involves measuring individual evaluations of crime seriousness and punishment severity; allowing individuals to assign punishment to crimes; and assessing whether the resulting match is explainable in terms of an equitable fit between the subjectively judged inputs and outcomes. Consensus on the norm can be measured at the two levels of principle and its application by a combination of tests within individual respondents and between respondents.

In order to obtain a model for just deserts, it is necessary to establish a theoretical and methodological framework. Social psychological equity theory provides a potential framework, as it is a general approach to social perceptions of justice within which some investigations of fairness of sanctioning have already been carried out. Examination of the equity theory approach to the present issue, however, proves to call for modifications in the previous account of the relation of equity to criminal

sanctioning as well as for modifications in the measurement techniques. In short, just deserts as a problem reveals problems in equity theory itself.

Equity Theory as a Framework for Just Deserts

Distributive justice is fundamentally describable as a logic of equivalence or balance between individuals. From Aristotle's Nicomachean Ethics to Adams (1965), this balance has been represented by an equation in which the input/output ratio of one party to an interaction should equal the input/output ratio of the other. Equity theory derives from Aristotle in basically attempting to evaluate the perceived fairness of different balances of contribution and outcome among individuals.

Recent attempts to extend equity theory to incidents of loss or harm--negative inputs by one party-- have occasioned both mathematical and conceptual confusion. In the two-party exchange, such concerns have necessitated modifying the mathematical treatment of fairness beyond the Aristotelian formula (cf. Walster, Berscheid, & Walster, 1973; Harris, 1976). Mathematically, the fundamental shift has been to replace a principle of proportionality as the central model with a principle of monotonic consistency between inputs and outputs for each party. Conceptually, the consideration of negative inputs has helped to spur the extension of equity theory into evaluations of fairness of exchange by third parties. The law itself can be seen as acting as a third party which evaluates the fairness of exchanges. Further, individual members of society can in turn evaluate how they feel the law should treat offenders. These assessments from individual citizens

can potentially tap the cognitive version of sanctioning norms.

However, recent attempts to apply equity theory to legal settings, summarized by Austin, Walster, and Utne (1976) and Walster et al. (1978), have not been entirely consistent and results have not been overwhelmingly strong. Why is this the case? The research in this area has involved the effects of pretrial suffering on subsequent evaluations of and sentencing of the defendant. As in prior equity research, the dominant concern has been how individual citizens would rectify inequity when it occurs. For example, Austin, Walster, and Utne (1976) investigated the effects on subsequent sentencing of a criminal's 'suffering in the act,' factorially manipulating this suffering as being at one of three levels (insufficient, appropriate, or overly severe). Across two crimes-- a purse-snatching and a mugging--the authors found that the subsequent sentence given to the defendant was lower if he had been subjected to pretrial suffering that was too severe.

The major potential determinant of sentencing, however, has gone uninvestigated in these studies. Legally and historically, the major overlay of equity upon sentencing would appear to be the notion that punishment should not be inordinate or arbitrarily linked to an offense. Even the Hebrew Lex Talionis, "an eye for an eye, a tooth for a tooth," incorporates this notion (cf. Biggs, 1955). Secondly, a further concern of legal systems has been equity among offenders. As Hart (1968) put this concern, the fair punishment system includes the condition that similarly situated defendants be treated similarly and differently situated defendants treated differently. Juries have historically responded both to the unfairness of

an available punishment being too harsh for the offense (Simon, 1967) and to the unfairness of one person receiving harsher treatment than another for a joint or commonly committed offense (Kalven & Zeisel, 1966). Structurally, criminal justice is set up as a balance between act and punishment, with the state as the third party judge, rather than as a balance between individual offenders and victims. The system's inputs are crimes and its outcomes are punishments. Its balancing act is thus between offenders across inputs and outcomes.⁴

If this model of criminal justice is appropriate, then the major determinant of sentencing should be looked for across offenders-- and it should be the offense committed. For example, the major fact leading to different sentencing should be whether one was a purse-snatcher or a mugger, not whether one suffered in the act of either purse-snatching or mugging. Extension of equity concerns to include evaluations by institutional third parties like the law needs to be accompanied by extension of the notion of what is being compared or balanced. The primary legal balance may be considerably abstracted from the specifics of the crime itself, and may involve other parties (i.e., other offenders) who form an implicit comparison group for any single case.

With certain modifications, however, the principle of just deserts in criminal sanctioning can be understood as a special case of the general principle of equity. It certainly involves inputs, outcomes, and the matching of the two to achieve supposed fairness. It simply extends and makes explicit what is implicit in all two-party assessments of equity: the evaluation of actual inputs and outcomes relative to other potential ones.

In raising the question of other potential inputs and outcomes, just deserts leads to the central measurement issue: how would-- or should-- one measure equity?

Measuring Equity

An equity principle is some formula for matching inputs to outcomes in a way which is perceived as fair. In most of the now-voluminous equity literature, this matching is expressed as an equation. Particularly since the advent of concern over handling negative inputs, there has been considerable debate over the correct form for such equations. However, this debate has been largely theoretical, as actual research generally assumes the equity principle to hold and tests the impact of various deviations from an equitable arrangement on the reactions of experimental subjects. Recent critiques (cf. Adams and Freedman, 1976; Harris, 1976) have suggested that the increasingly elaborate equity formulae appear to call for different forms of research design and/or levels of measurement from those previously employed-- in short, that equity as a principle or formula deserves to be tested.

What would it look like to test an equity formula? Examination of the existing literature suggests that two issues have been hidden by the prior interests of researchers and the prior strategies of research. First, is equity genuinely used by individual respondents to order their perceptions of the world? Second, do individuals consensually agree on what is an equitable outcome given a specific input? These are of course specific forms of the general issues suggested earlier as being central to the testing of any norm as a cognitive structure.

The typical equity research design is a factorial analysis of variance in which experimental subjects each receive one experimental treatment. Thus in

the Austin et al. (1976) study, subjects read about an offender who received one of three levels of injury in the course of a crime. In such studies the evaluation of the inputs as high/medium/low is typically arrived at either by pretesting to obtain consensually agreed upon stimuli or by recourse to common sense. The equity principle is then used to predict the experimental outcome, which is counted as confirmed by the presence of significant differences between average outcomes assigned in the various experimental conditions. Conversely, the outcomes can be specified and evaluations of the inputs obtained. Thirdly, both outcomes and inputs can be described and subjects asked whether the arrangement is fair. Such designs all share three properties: first, they implicitly or explicitly fix two components of the formula and test the third; second, they employ essentially ordinal level comparisons between levels of treatment conditions; third, they either assume or establish a priori consensus on stimuli. None of these three properties is necessary for the testing of equity as a theory. In fact, all of these three properties must be altered for the full-blown testing of equity.

The typical research design can be contrasted with one which actually tests whether equity holds at the level of the individual respondent. This alternative design has been alluded to briefly in the discussions of normative consensus and just deserts; now it can be fleshed out in terms of its contrast to prior equity research. The requisite ingredients are the individuals' matches of objective inputs to outcomes; their subjective assessments of these inputs and outcomes; and a theoretically plausible equity model with which to assess the matching. For example, the logical first step in testing the distributive justice norm "equal pay for equal work" would be to get the

respondent to say which work inputs should go with which pay outcomes.⁵ This is the respondent's own version of justice. But each objective stimulus has a potentially unique subjective meaning to the respondent. If the respondent is answering equitably according to theory-- if equity indeed lies in the eye of the individual beholder-- he or she should be matching objective inputs to outcomes consistently with their subjective meaning. Thus if one can obtain the respondent's own subjective assessments of the value of inputs relative to other inputs and the value of outcomes relative to other outcomes, one can test whether the respondent's matching is internally equitable by substituting the subjective ratings for the objective stimuli. In such a design, each respondent effectively serves as an experiment in model-fitting. The equity model lies in the eye of the researcher rather than the beholder; researchers can therefore say whether the respondent's version of "equal pay for equal work" fits the theoretically predicted version, but cannot say what other principles, if any, the respondent is bringing to bear on the task.

In testing whether the punishment should fit the crime, an analogous model-fitting operation can be used. Finding out which objective crimes are assigned which objective punishments does not test equity at the individual level. One needs to know how serious the respondent considers the crime and how severe the punishment. Then the respondent's matches of a series of crimes and punishments can be tested against a model which predicts their matches from their subjective ratings of seriousness and severity. Just as "equal pay for equal work" is shorthand for "the value of outcomes should match the value of inputs," the notion that the punishment should fit the crime is shorthand for consistency between the seriousness of a crime and the severity of the punishment it receives.

Such a model-fitting design also allows the researcher to test consensus rather than setting it a priori, and to test two distinct versions of consensus. Consensus on the justice principle involved is represented by high level of individual fitting of the specified model. Consensus on specific social stimuli is represented by a high level of between-individual agreement. Potential differences between these two forms of consensus can also be assessed, as can demographic bases of disagreement. A within-respondent design makes it possible to pull apart normative consensus into its component parts.

The present study uses this proposed alternative design for testing the norm of fitting the punishment to the crime. The study deviates from prior equity research both substantively and methodologically. Substantively, we argue that the central principle to be assessed in the arena of criminal justice is just deserts, the notion that the punishment should fit the crime. Methodologically, we argue that the appropriate test of equity (as of any cognitively held norm of justice) involves an interplay between individual adherence to principle and individuals' agreement on facts. The assessment of consensus on such norms calls for consideration of both the principle and its application. Given the appropriate within-respondent design, both cognitive principles of fairness and social consensus on fairness can be assessed simultaneously.

Methods

Scaling Techniques: Providing Ratio-level Judgments of Crime and Punishment

Human judgments of the intensity of many physical stimuli can be shown to fit a power function such that the relationship between stimulation and

sensation is described by the equation

$$y = cx^n \quad (1)$$

(Stevens, 1975). Stevens' Law, as this relationship is known in the psychophysics literature, translates into a rather simple verbal conception: that equal stimulus ratios produce equal subjective ratios. Thus the perceived intensity of many physical attributes can be ratio scaled. Recently, psychophysical scaling techniques have been extended into measurement of the intensity dimension of social and political attitudes and attributes-- such as occupational prestige (Kunnapas and Wilkstrom, 1963), social status (Hamblin, 1971; Rainwater, 1971), importance of political offices (Shinn, 1969), support for political institutions (Lodge, et al., 1976), and, of course, seriousness of crimes (Sellin and Wolfgang, 1964).⁶

Psychophysical scaling of social variables offers a highly developed methodology backed by rigorous theoretical predictions concerning physical modalities. It has two major advantages over category and JND (Just-Noticeable-Difference) scales such as Likert and Thurstone scales: it places fewer constraints on respondents, and it provides established procedures for validation using the technique of cross-modality matching (see footnote 6 and Cliff, 1973). In addition, two modalities used extensively in psychophysics prove to be adaptable to survey investigation. Magnitude estimation (ME), simple assessment of intensity by assigning numbers, and line production (LP); assessment of intensity by drawing lines of various lengths, have convenient mathematical properties (cf. Hamilton and Rotkin, 1978). In addition, it is quite feasible to get survey respondents to give number estimates for how many times more intense one social stimulus is relative to another; indeed, ME was the method

used by Sellin and Wolfgang (1964). It is somewhat more cumbersome, but also practicable, to get respondents to draw lines of various lengths to express their assessments of relative intensity. (Other possible modalities such as adjusting the brightness of lights or squeezes of a hand would require equipment such as a rheostat device or a hand dynamometer which are hardly in the typical interviewer's experience and which might turn the interviewer into a somewhat formidable portable laboratory technician.) Thus the most mathematically convenient modalities are also the most physically convenient ones, and render the survey instrument a potential place to attempt psychophysical scaling of social stimuli.

The present investigation uses a data set of psychophysically derived scales produced for another purpose. In a research program on public opinion and capital punishment, psychophysically validated scales for seriousness of crime and severity of punishment were developed (see Hamilton and Rotkin, 1976, 1978). In the data set used here, respondents were additionally asked to indicate what punishments they would assign to the crimes being scaled. Given the purpose of the research program, the scales and the data derived from them are not optimal for present purposes. The essential problems are a crime set and punishment set which emphasize potentially capital offenses and relatively severe punishments, providing scales with attenuated range. Happily, from the point of view of scientific inference, these difficulties make the data set a relatively stringent test of current concerns. Relationships found here both at the individual level and at the level of aggregate consensus are likely to be low estimates of actual relationships.

A couple of features unique to psychophysical data make the variables somewhat unusual. First, the data are theoretically log-normal, since errors in

ratio judgments are multiplicative errors. Secondly, respondents are allowed to pick an arbitrary standard or "modulus" against which to assess relative seriousness and severity.⁷ In keeping with typical handling of such data in psychophysics, the measures used in the analyses are not raw scores. Instead, they are averages of the ME and LP judgments for a given crime or punishment. These are logged and their mean is subtracted (equivalent to dividing the raw scores by geometric means) to remove the effect of the subject's modulus. In psychophysics, individual level scores are typically not employed at all; group averages are used instead. Here, in order to test just deserts as a cognitive principle, the scores for individual respondents are employed.

The Data Set: The Boston Survey Sample and Interview Instrument

In February and March, 1976, a quota sample of 391 respondents from the Boston SMSA was obtained. There was no particular interest in generalizing the results to Boston per se, but merely in obtaining responses from adults with a variety of demographic characteristics. In addition, prior research by Sellin and Wolfgang (1964) using differing purposive samples, by Rosal et al. (1974) using a quota sample of Baltimore; and by other investigators replicating the Sellin-Wolfgang scale in a number of different locations and cultural groups, had shown crime seriousness to be a remarkably consensual and stable judgment. Thus it was felt justifiable to use quota sampling rather than strict probability sampling to gain the significant budget savings possible with the former. All sampling and interviewing were done by the Survey Research Program, University of Massachusetts, Boston.⁸

The data-set consisted of a face-to-face interview of approximately 45 minutes duration and a dropoff questionnaire with an additional approximately 45

minutes of material. In each part, roughly half of the survey material consisted of psychophysical scaling judgments; the remainder of each part contained attitude items related to crimes involving authority (e.g., My Lai, Watergate) irrelevant to the present investigation. Respondents were paid \$5.00 for mailing back the dropoff questionnaire. In all, 314 (81%) did so, most without even a follow-up telephone call. Material from the dropoff questionnaire forms the data for the present analyses.

In the interview, respondents first completed a training task: matching subjective magnitudes of lines to stimulus numbers and numbers to stimulus lines. This task served both to familiarize respondents with psychophysical judgments and to provide data for technical corrections.⁹ Respondents then rated the seriousness of the 17 crimes in the crime seriousness scale, first by making LP judgments and second by making ME judgments. For both sets of judgments the stimuli were arranged in random order, with each set beginning with a crime of intermediate seriousness. Similar instructions and procedure were then used for rating the 10 punishments. Items making up the two scales are reproduced below in Table 1.

Insert Table 1 About Here

Respondents were asked to do two tasks in the dropoff questionnaire. First, the crimes were presented in random order, and respondents were asked to assign each crime an appropriate punishment, using the list of the punishment scale items. The punishment items were arrayed in the order of average severity established in a pretest and were assigned letters A through J. Respondents were encouraged to write in a response if no punishment on the list seemed appropriate. This task provides data for the matching of an abstractly defined

crime and an accompanying punishment, since the respondent's own psychophysical scores from the main instrument could be assigned to each crime and each punishment. Second, respondents were asked to judge concrete examples of crimes, written as brief vignettes. The total of 14 vignette crimes included an even heavier mix of serious crimes, as they were used to obtain judgments of several actual capital cases. They also excluded some of the abstract crimes for which it was difficult to obtain examples of typical offenses. (Other than the actual capital cases, vignette examples were actual "typical" cases provided by a consulting lawyer.) Respondents were asked both to assign psychophysical seriousness scores to the vignettes using LP judgments and to assign punishments from a list as they had done for the abstract crime labels. These vignettes thus provide data on the match between a concretely described crime and its punishment, using the LP judgments provided in the dropoff instrument for seriousness and substituting the psychophysical values of the punishment judgments made in the main interview for severity. For present purposes the abstract and vignette tasks provide an internal replication on fitting the punishment to the crime at different levels of concreteness.

Armed with psychophysical scales for crime seriousness and punishment severity, and with judgments from respondents of the punishments appropriate for these crimes, how would one assess the degree to which just deserts is a principle followed consistently by individuals or a practice followed consistently across individuals? Such an assessment requires a formalization of the equity principle at the appropriate level of measurement.

Formalizing the Principle of Just Deserts for Tests of Equity and Consensus

Fortunately, with the notion of just deserts we can avoid the issue of positive versus negative inputs that has so far plagued the development of single equity equations (cf. Harris, 1976). In considering seriousness of crime and severity of punishment, all inputs are of the same logical type; the only issue is the degree of seriousness, not whether some deeds are good and others bad. Given this simplification, and the fact that just deserts appears to imply proportional matching of crime seriousness and punishment severity, the fundamental equation of just deserts is the following:

$$\frac{P_1}{C_1} = \frac{P_2}{C_2} \quad (2)$$

Since equity theory is an individual level theory, this equation should be tested for all persons *i* and for all crimes and punishments *j*.

Equation (2) can be rewritten as:

$$P_{i,j} = k_i C_{i,j} \text{ for all } i,j \quad (3)$$

Since our data are log-normal, the logged equation is the appropriate one to be estimated:

$$\log P_{i,j} = \log k_i + \log C_{i,j} \quad (4)$$

This equation, in turn, implicitly assumes that the slope of *C* on *P* is one; in a regression analysis, this is an easily estimable parameter. The logged equation then becomes a familiar regression model:

$$\log P_{i,j} = \log k_i + b_i \log C_{i,j} \quad (5)$$

Transformed back from the logs, this becomes

$$P_{i,j} = k_i C_{i,j}^{b_i} \quad (6)$$

Equation (6), of course, is simply an individual-level expression of the power function. Expression of the equity equation in the logged form with ratio data makes it plausible to restate equity as a power law, and to consider slopes of inputs on outputs that are other than unity.

Equation 5 is the estimable specification of the equity principle as a relation between proportional judgments made by an individual. As a summary of these judgments, the model is potent to the extent that respondents' evaluations of matched pairs of crimes and punishments can be described by this linear relationship in the logarithms. If the model accurately describes individual judgments, as shown by high correlation between crime seriousness and punishment severity, then the equity principle (as operationalized) accounts for respondents' answers. Of course, it does not follow that respondents need have had the equity principle consciously in mind in order to produce a good fit to the model; but good fit does mean that the analyst can at least conclude that respondents' behavior is consistent with having applied the principle. Equation 5 can be estimated twice for each respondent-- once for the abstract crime labels and once for the specific vignette examples. For each respondent there will be two estimates of how closely their responses conform to this operationalization of the just deserts model.

In fitting these equations, other summary statistics-- the slope and the intercept-- become available. The slope captures the subjective proportionality of justice: i.e., the scaled increase in punishment severity that accompanies a unit increase in crime seriousness.¹⁰ (See footnote 10 for the interpretation of this coefficient in the antilog form, equation 6.) The intercept is the average "profits" (or losses in the negative domain of punishments) assigned. Thus two distinct senses of equity can be invoked. The output (punishment) can

be fitted to the input (crime) in the sense of proportionality: i.e., any unit increase in the seriousness of the crime is matched with a scaled increase in the severity of the punishment. In addition, the typical net outcome-- the intercept-- can be at issue. Even if proportionality holds, it may not seem fair to punish the theft of \$50 with life imprisonment and more serious crimes with proportionally graded degrees of torture and humiliation. Both the slope and the intercept in this formulation can have interesting equity interpretations.

To our knowledge, estimates of the actual coefficients in an equity equation have not previously been made. Indeed, such estimation would not be feasible in the typical factorial design in which inputs and outcomes are ordinarily distinguished into categories such as high versus low (cf. critique by Adams and Freedman, 1976). As Adams and Freedman suggest, these actual coefficients could be of interest. Further, since two estimates are available for each respondent, the reliability of these components of the idiosyncratic equity equation can be ascertained.

Overall, there are two sets of tests of the equity principle as a within-individual pattern. First, the within-individual consistency of the slope and intercept terms can be examined to see if stable idiosyncracies characterize respondents' matched evaluations of seriousness and severity. This analysis can also be interpreted as testing the reliability of the slope and intercept. But the more important issue is the potency of the principle itself in describing responses. A simple measure of the degree to which the data fit the principle is the correlation of crime seriousness to punishment severity within individuals. This statistic indicates whether people follow the principle of fitting the punishment to the crime, and further provides a continuous measure of the degree to which they do so.

There are, however, two separable notions of normative consensus: use of the cognitive principle and agreement on the specific stimuli involved. The tests for the two are logically distinct. Respondents might agree that increases in crime seriousness should be matched by increases in punishment severity without agreeing on which actual crimes are the more serious, which punishments are more severe, or which crimes should get which punishments. Evaluations of consensus within and between individuals involve different tests.

One measure of consensus between individuals is the ecological correlation of the average seriousness rating of crimes with the average severities of the punishments they receive. Such correlations are used as indices of consensus in literatures as diverse as crime seriousness and occupational prestige (cf. Hamblin, 1971; Hodge, Siegel and Rossi, 1964; Sellin and Wolfgang, 1964). In a sense, the factorial designs of prior equity experiments can also be seen as exercises in ecological correlation, given that hypotheses are taken as supported when different stimuli (implicitly assumed to be perceived the same by all respondents) produce significantly different average responses. Yet it is well known that it is invalid to infer individual level relations from aggregated data (Robinson, 1950). In this case such inference would be that of individual level conformity with the principle from the correlation of the group averages. It is equally invalid to infer aggregate agreement from individual level conformity, since there may be between-person disagreement on the specifics as noted above.

It is thus useful first to ascertain the ecological level correlation and then to assess its relation to individual level fit of punishment to crime. Individual deviations from group consensus on either crime seriousness

or punishment severity may be essentially irrelevant, but it is important to test whether individuals' own deviant perceptions or ratings are necessary parts of a model predicting their responses from the group averages. If individual perceptions are unnecessary, then one of two conditions must be the case: either individuals magically all learn about crimes and punishments so as to evaluate them identically while they go about using the equity principle, or apparent individual use of the equity principle is actually the spurious outgrowth of conformity to a social consensus about certain stimulus objects and their matching. Thus evaluating the relation between aggregate ratings and individual scores is crucial to demonstrating whether both consensus on social facts and consensus on the principle actually exist.

More importantly, evaluating the relation between aggregate and individual ratings reveals a potential conflict between the two notions of consensus, between conformity to the equity principle and agreement with the aggregate. To the extent that the individual respondent holds to the equity principle, any deviance in evaluation of a crime's seriousness will be matched by a correspondingly deviant assessment of the appropriate severity of punishment. In other words, consensus on a principle at the cognitive level leads to principled dissent from the group when any difference in evaluating stimuli is encountered. A simple structural equation model will be used to test for the presence and magnitude of principled dissent accounted for by adherence to the principle of just deserts.

A final issue is the possible presence of group-based disagreement on either the just deserts principle itself or the social stimuli presented. Failure to follow the principle can be either random, unrelated to any group cleavages, or systematically linked to demographic categories. From the

standpoint of a conflict versus consensus model of criminal justice, the most interesting variables are those indicative of hierarchical status in the society-- race, income, and education. If failure to fit the principle is associated with such cleavages, it lends itself particularly well to a conflict interpretation, given the known status differences in likelihood of entering the criminal justice system as a defendant. In addition, examination of the degree of agreement with the group averages by different demographic groups provides an indication of whether any groups disagree about the social facts as well as the cognitive principle. Thus the two versions of normative consensus-- agreement on principle and agreement on stimuli-- can be examined for the presence of hierarchically related disagreement.

First, however, a simple question: in the eye of the individual respondent, should the punishment fit the crime?

Results

Equity at the Individual Level

The most stringent version of the current model, equity as an individual-level power law, involves the betas for each respondent. The slope of punishment on crime in the logged equation-- the exponent for crime in the power function-- need not be 1.0, as has been noted, for a law-like version of matching the two to hold. Thus one issue is the average sizes of the individual betas in the two measurements, those for abstract crime labels and for specific vignettes. A second issue is the range of betas in each. Respondents might show wide ranges in the slope of punishment on crime, which would indicate that the aggregated equation would not adequately describe the individual-level "law." The size of the betas as such has no clean interpretation.

as individual differences can be a function of shrinking or stretching either scale. For example, a high beta could reflect either a stretched-out punishment severity scale or a narrow range in the crime seriousness scale. If different sized betas reflected such shrinking and stretching, indicative of different power laws for different respondents, one would expect this to be a stable difference between persons across tasks. Even with very different betas, then, respondents could each be following a personal, specific power law. This possibility can be tested by examining the correlation between each person's betas for the abstract and vignette tasks. A high correlation would indicate a consistent individual-level matching of crime to punishment, irrespective of differences in the exponents that describe the idiosyncratic matching functions.

Table 2 presents the relevant tests for size and consistency of the betas, as well as the alphas and r 's to which we will turn next.¹¹ The data show a close one-to-one match on the average: the average betas from both tasks are quite close to 1.0. However, substantial individual variability around that average is found in both abstract and vignette measures. Do individuals then each fit a personalized "law" or matching function utilizing their own personal exponent (beta)? As the significant but relatively small correlation between the two betas indicates, this does not appear to be the case. Treating the correlation as a test-retest or reliability measure, an r of .25 between the scores is too low to indicate that individuals have a stable pattern which they apply to both tasks. Thus the results argue against a stringent individual-level power function formulation of just deserts.

Insert Table 2 About Here

The alphas present a different picture at this level of analysis. Although the intercepts need not be zero, they proved to be quite small in both abstract and vignette tasks, with lower variability than the betas. These average intercepts indicate that a small weighting constant is necessary in the power function fit of crime to punishment for most respondents. The alphas, which reflect the average punishment employed by respondents, are also substantially more stable across the two tasks. Further evidences of the reliability of alpha as an indicator of average punitiveness were substantial correlations between alpha and average punishment severity (abstract: $r = .88$; vignette: $r = .68$. Both tests $p < .001$) and between alpha and an average ordinal measure of punishment severity (abstract: $r = .72$; vignette: $r = .55$. Both tests, $p < .001$). Thus respondents' intercepts would appear to hold promise as an indicator of individual punitiveness, an issue to which we shall turn later.

The central and simplest test of whether the respondents fit the punishment to the crime remains to be examined, however. Even if the unreliability of the betas forces rejection of the notion of personal, specific power laws, a high correlation between crime seriousness and punishment severity would indicate that individuals are following the equity model with the requisite monotonicity suggested by Harris (1976). As Table 2 indicates, the correlation between crime seriousness and assigned punishment severity is substantial for both abstract and vignette tasks. There appears to be a strong tendency toward a monotonic match between the judged seriousness of a crime and the severity of the punishment one assigns to it. Median within-individual correlations of such size are striking indeed.

From one point of view, this tendency to fit the punishment to the crime in a monotonic way is not a stable property of individuals across tasks. The correlation between the two different correlations produced by each individual was only .33. (The correlation ratios were normalized by a Fisher z-transform (McNemar, 1969, p. 157) before this correlation was calculated.) The quality of the fit of punishment to crime on one task is not a very reliable predictor of the fit on the other task. Yet most of the discrepancies between correlations were modest, and the tendency to a strong fit was consistent in broad pattern if not in detail; 82% of the respondents who completed both tasks had correlations of crime seriousness with punishment severity of greater than .5 on both tasks. The scatter plot of the two correlations showed mainly a large "ball" centered in the sector where both correlations were greater than .5.

Although the use of correlation coefficients to evaluate the crime-punishment fit leads to discarding some of the information from the ratio scaling tasks, it does indicate that a principle of monotonicity describes quite well most people's response to the tasks. Within the abstract and vignette sets of stimuli, knowledge of one set of judgments, on either the seriousness ratings or the severities of assigned punishments, tells a great deal about the other set of ratings. The summary standard of good fit, both correlations above .5, is arbitrary and includes as fitters some people whose correlations were far short of perfect linear association. But we think it is a fair summary of these data that most people who performed the task gave responses that were quite congruent with a monotonic version of the principle of just deserts, that the punishment should fit the crime.

Given that individual respondents demonstrate a striking degree of structure in their fit of punishment to crime, we can then turn to the second

aspect of normative consensus: agreement between individuals on the concrete stimuli. Individuals adhere to the principle, albeit not at the sophisticated level of a power law. Do they also agree on the facts?

Social Consensus on Crimes and Punishments

The typical indicator of consensus in much of the literature would be the ecological correlation between crime seriousness and punishment severity. In these data, the correlations between the average seriousness ratings for each crime and the average severities of punishments assigned are quite high. For the abstract crimes, with an n of 17 (for the 17 crimes), the group-level correlation was .98 ($t = 18.191$, $p < .0001$). For the vignette crimes, with an n of 14, the corresponding figure was .97 ($t = 15.354$, $p < .001$). Thus the average seriousness assigned to a crime was closely congruent with the average severity of the punishment it was given.

This appears to be a powerful preliminary indication of consensus between individuals. The question of consensus might even be rephrased from this point of view to ask whether individual differences matter at all, or whether we can simply predict any individual's punishment assignment on any offense adequately from knowing the group mean for the seriousness of that offense. An even more radical question can also be raised: do we really need the concept of crime seriousness as an "input," or is crime seriousness simply a surrogate measure of the average punishment assigned?¹² Is this relationship so strong that it must represent two measures of the same thing?

The investigation of these issues in consensus between subjects required some modifications in treatment of the data set. For the first time, the fact that multiple measures were obtained from each respondent became a problem:

thousands of crime and punishment judgments were made, but the number of respondents was much smaller. As a first pass through this portion of the data analysis, artificial between-subjects data sets were created.¹³ In the case of the 17 abstract crimes, 17 data sets were formed; in each data set, a given respondent was represented only once by means of a random selection criterion. Within each data set, the repetition of measures was not relevant. The 17 data sets had slightly different n 's because of missing data, but averaged over 280 cases with roughly equal representation for each crime. The same procedure for the 14 vignette stimuli yielded 14 data sets of similar size and composition. Across all analyses performed, the multiple data sets showed such remarkable consistency in results that they were pooled into a single large data set for presentation purposes. In effect, there is either a single data set with thousands of cases for the abstract crimes, as well as one for the vignettes, or there are 17 quasi-independent replications of the same finding for the abstract crimes and 14 for the vignettes. Although we would be hesitant to make significance tests of results based on the two large data sets, the magnitude of observed effects made such tests superfluous. Overall, the stability of findings warranted the simplified presentation of results.

Can seriousness ratings simply be explained away? As a first step in answering this question, we regressed both individual seriousness ratings and punishment ratings on dummies for the crime stimuli, to ascertain how much of the variation in either rating was explained simply by knowing the stimuli. In the abstract crime set, with a case base of 4,595, the regression of punishment on the 16 dummies for stimuli produced a multiple $r = .62$ ($R^2 = .39$); the regression of seriousness on the dummies yielded a multiple

$r = .67$ ($R^2 = .45$). Similar relations held in the vignettes, with a case base of 3,742: the multiple r for punishment was $.63$ ($R^2 = .40$), and the multiple r for seriousness was $.72$ ($R^2 = .52$). Thus a great deal of the individual variation in both seriousness and punishment ratings, can be accounted for simply by knowing the stimulus.

Does knowing respondents' individual seriousness judgments really say anything more about the punishment assigned than just knowing what the stimulus was? Table 3 summarizes the subsequent analyses addressed to this issue. The logic involved three steps, each the same (and with similar results) in the two sub-data sets. The first step was to ask the relation of respondents' punishment assignment to seriousness assignment in each data set. In the abstract crimes, this produced a somewhat lower r and R^2 than did the simple dummies. In the vignette crimes, this analysis produced results quite similar to those for the dummies. In both data sets, the second step was to ask whether knowing the individual's own seriousness rating significantly adds to the variance explained by the dummies alone. Despite the somewhat worrisome performance of seriousness per se among the abstract crimes in the previous analysis, the R^2 was substantially increased for both abstract and vignette crimes by considering both the stimulus and the respondent's unique seriousness ratings.

Insert Table 3 About Here

The final step in Table 3 represents a twist on the multicollinearity observed. If stimuli and seriousness ratings are so highly related, then we can explore a final substitution: replacing the dummies for crime stimulus by the group average seriousness and the individual's own seriousness ratings-- which, when in the same equation with group average, becomes the individual's

deviation from the average. Results for this equation strongly resemble the previous one and reconfirm the significant role of individual deviation on the specifics of evaluations.

These results indicate that we can, indeed, gain something by using the notion of the seriousness of the crime as an input to its assigned punishment-- that seriousness is not the same thing as punishment assigned, although it correlates highly with it. Further, both group consensus and individual deviation from consensus on seriousness are significant predictors of individual punishment assignments. No further steps are needed. Tests for whether the interaction between the crime stimulus and its seriousness was needed in the model yielded increments of less than 1% in R^2 in both abstract and vignette crimes. Thus the analysis of consensus between individuals reveals both substantial consensus and some room for effect of individual idiosyncrasy.

Although these analyses have confirmed the importance of both group averages and individual idiosyncrasy as determinants of the individual's response, they have not clarified the extent to which individual responses correspond to agreement with the group average. To do so involves still further steps in pulling apart the ecological correlation and relating it to apparent use of the principle at the individual level. These steps reveal the extent to which the remarkable structure already found in within-individual fitting of the equity model can be accounted for by the remarkable agreement between individuals on the stimuli.

Decomposing Agreement: Applying the Principle Versus Agreeing with the Group

An individual's apparent use of a principle like just deserts can conceivably represent nothing more than rote learning of how things are usually evaluated and paired up in a society. For example, the socially understood

seriousness and punishments for various crimes may be learned by rote-- perhaps through a process of repeated association, on TV, in the newspapers, and so forth. Under this model the group averages reflect the outcome of this learning process and individual ratings are imperfect reflections of these aggregate social facts. The extent to which individual responses display an apparent pattern of just deserts then would not reflect their personal adherence to an equity principle. Instead the individual level correlations would be a joint result of the extent to which the group averages displayed a monotonic fit and the degree of success of the individual in reproducing the group averages.

The ecological correlations of .98 and .97 certainly demonstrate a monotonic relation between aggregate seriousness of crime and severity of its punishment. The success of individuals at reproducing group averages also proves to be high for both stimulus sets. The relevant correlations are those between individuals' own seriousness scores and the group average and between individuals' punishment severity scores and the group average. (Similar scores have in fact been used as measures of normative consensus in prior research, notably by Rossi *et al.*, 1974.) For the abstract crimes, the average correlation between an individual's seriousness score and the group mean score was .71; the average correlation between punishment severity and the group mean score was .73. For the vignette crimes the corresponding average correlations were .77 and .75. Thus both conditions for the apparent production of "rule" learning by "rote" learning would appear to be met.

What would such a model formally look like? The rote learning version can be modeled with a structure like that in Figure 1, Model 1(a). Under the assumptions of this model, an individual's seriousness ratings are some reflection of the aggregate seriousness ratings. Their punishment ratings

are a reflection of the aggregate punishment ratings. Since the fitting of individual seriousness to aggregate punishment and the fitting of individual punishment to aggregate seriousness would require the application of an individual equity principle, these causal paths are excluded. Under these assumptions, the individual level fit of respondents' seriousness judgments to their punishment judgments can be found by multiplying around the model (Duncan, 1976). It would therefore equal the product of the individuals' fit of seriousness to aggregate seriousness, the aggregate fit of seriousness to punishment, and the individuals' fit of punishment to aggregate punishment.

Insert Figure 1 About Here

In Model 1(a), arrows have also been drawn to represent the correlation of errors. These errors are the individuals' deviations from the aggregate seriousness and punishment ratings. Uncorrelated deviations would indicate that the fit of the punishment to the crime at the individual level can be accounted for by agreement with the group averages. The (positive) magnitude of the correlation of deviations indicates the extent to which the equity principle is reflected in the individuals' deviations: in other words, a respondent who was deviant on seriousness ratings should tend to be deviant in a consistent way on punishment assignments. The correlation of errors represents the effect of applying equity at the individual level after we have partialled out that portion of the fit that can be attributed to consensus. It is a conservative test for the extent of use of the principle, given that a principled rule applying respondent who coincidentally agreed with the group about the stimuli would not show such correlated errors. However, the correlation of the deviations from the model is unambiguously a reflection of the individual application of the principle, while the portion of fit

attributable in this model to consensual forces could reflect simple rote learning rather than rule learning. If equity actually lies in the eye of the individual beholder, it can be found through a conservative test by the search for correlated errors.

Model 1(a) is not always appropriate for all persons because its assumptions imply two identification restrictions. When these restrictions do not hold exactly, as they often do not with the small number of cases used to estimate each of the 594 applications of the model, estimates of the error correlation that exceed the boundaries of the interval [-1,+1] can occur. For that reason, we actually estimated Model 1(b) to measure the correlation of the deviations. The argument that correlation of deviations is an estimate of principle adherence that cannot be attributed to rote learning still applies for this just identified model.

Table 4 presents the relevant statistics on the extent of correlated error for each set of crime stimuli. The average correlations of errors actually mask the extent of such error, as they are affected by skewed scores. Thus the summaries in Table 4 of the median correlated errors and of the percent of respondents whose errors are positively correlated are more appropriate indices of use of the equity principle over and above agreement with the group. Despite the modest size of the average correlations, results clearly indicate that most respondents' errors were positive ones: when respondents deviated from the group on the seriousness of an offense they tended to deviate consistently on the severity of the punishment they assigned it.

Insert Table 4 About Here

This conclusion-- the finding of principle over and above consensus on fact-- hides a sharp edge. Adherence to the principle that the punishment

fit the crime is actually a form of normative consensus that is quite imprecise, as there is considerable latitude left for disagreement about which punishment should be fit to which crime. Further, given any social disagreement about judgments of the specific stimuli involved, a tension emerges: the greater the adherence to principle, the more systematic the deviation from the group. Consensus on principle implies principled dissent on application.

The two stimulus sets, the abstract and vignette crimes, represent an interesting contrast in this regard. As Table 4 indicates, the average correlation of errors is larger in the vignette crimes. This is mathematically inevitable, given that the degree of agreement with the group about both crimes and punishments was similar in the two data sets, as noted above, and that in the vignettes individual adherence to the principle itself was higher (as presented in Table 2). The finding simply demonstrates empirically the point that greater adherence to principle can lead to more systematic deviation from the group. But it is conceptually interesting to speculate on why the vignettes are different. Because there are methodological differences between the procedures for the two stimulus sets, as described in the methods discussion, we cannot be completely confident about explaining this difference substantively. However, we are inclined to believe that the greater use of the equity principle in the vignette crimes reflects their concrete quality: that respondents faced with a concrete case are better able or more willing to link up their assessment of its seriousness with the punishment they mete out in a way consistent with just deserts. If this interpretation is correct, then the closer things get down to cases-- of real-world complexity and detail-- the greater the incidence and degree of principled dissent is likely to be.

Principled dissent is an intriguing notion, but it still represents basic adherence to the abstract norm, if not full agreement with the group. Thus a final probing must be made. Is there any evidence in a data set characterized by such large correlations that would point to what is normally thought of as a conflict rather than a consensus? In other words, are there any social groups who fail to use the principle, fail to agree with the aggregate on stimuli, or both?

Not Playing the Game: Demographics and Just Deserts

For a conflict interpretation of disagreement about just deserts, the central demographic variables of interest are those indicative of hierarchical standing in the society: race, income, education, and possibly sex. The central issues are whether members of any demographic categories are more or less likely either to fit the punishment to the crime or to agree more closely with the average responses to specific stimuli. Therefore the focus shifts to variations in the relevant within-individual correlations: those that variously reflect fulfilling the principle and those that reflect the similarity of individuals' judgments of seriousness and severity to the average judgments.¹⁴

Table 5 presents the zero-order correlations of relevant demographics with a series of variables indicative of fitting the punishment to the crime, as well as with two indicators of punitiveness (alpha and average punishment severity). All correlations were normalized by Fisher's z -transform before being treated as dependent variables in relation to the demographics, including the partial r from Figure 1, Model 1(b).¹⁵ (See Anderson, 1958, for the z -transform for a partial correlation.) Where correlations were nonsignificant, they are presented in smaller type in order to facilitate visual presentation of the overall pattern of results that emerged.

The pattern is weak but clear. Across all measures, black respondents are less likely to fit. They are less likely to show a high fit of the punishment to the crime, less likely to agree with the aggregate on seriousness judgments, less likely to show a high alpha or average punishment assignment.¹⁶ Across all measures, high income respondents are more likely to fit. They show high fit of the punishment to the crime and high levels of agreement with the aggregate. We should stress that no results are of very impressive size, particularly in contrast to the sizes of the consensual components already presented. However, the patterns for race and income are at least consistent if not powerful.

Only three specific findings were significant across both data sets, a standard which would appear to be a reasonable criterion for believing any particular piece of the pattern. These results were the black tendency to disagree with the aggregate seriousness ratings, the high income tendency to agree with the aggregate punishment ratings, and a surprising tendency for males to show higher levels of principled dissent (as indexed by their partial r). Since no other effects for sex were consistently significant, and some were even inconsistent with respect to sign, we are left with no particular clues concerning the meaning of the results for sex.

One set of effects is fortunately missing. The absence of any significant effects for education makes the interpretation of results for the race and income variables more clear. Since all three variables are related, the simultaneous presence of either of the other effects with education effects could be interpreted in terms of such issues as knowledge or sophistication in handling a complex questionnaire task. These interpretations do not appear viable; nor did regression analyses including all the relevant

demographics make them any more so, as education effects continued to be absent. Thus we believe that it is reasonable to assume that low income and black respondents knew what was being asked but disagreed with it systematically. It further appears that some of their lack of fitting the punishment to the crime results from disagreement about crime seriousness (in the case of blacks) or punishment severity (in the case of low income respondents). Both divergence on use of the principle and divergence on perception of the stimuli are evident, and both appear to be real disagreements rather than functions of sophistication.

Consistency in the fit of the punishment to the crime across the two tasks can also be presented in a cross-tabular form for ease of visualization. In order to do so, a categorical dependent variable was constructed to reflect whether one hit the criterion of .5 crime-punishment correlation on both tasks (82% of the respondents) or not. Table 6 presents the zero-order cross-tabulations of bad versus good fit with race and income, the only significant predictors of good fit by this measure. Such tables are essentially heuristic devices, as the cut point for correlations is obviously arbitrary and the cut points for income are somewhat arbitrary reflections of the marginal distributions. Nevertheless, the relationships of each demographic to the overall criterion of good fit are vividly illustrated in tabular form: to be black or low income means that one is likely to fall outside of the group of "good respondents" who faithfully match the punishment to the crime. Higher-order tabulation showed that among white respondents income continued to have a significant effect; the small number of blacks made higher-order tabulation inappropriate.

Insert Table 6 About Here

Overall, then, disagreement on fitting the punishment to the crime is related to hierarchical social status, specifically to income and race. Despite the broad general tendency of respondents to fit the model, low income and black respondents show significant disagreement. Yet so far a more subtle possible form of disagreement has been ignored: not filling in punishment responses on some of one's crime stimuli. Only a few respondents were missing an entire task; many respondents, however, availed themselves of the option of writing in a response on one or more single crimes. For the least serious crime-- taking \$50-- fully 28% of the sample wrote in other answers for both abstract and vignette versions of the crime. On most such write-ins, respondents favored a punishment of less than two years, which anchored the provided scale. These write-ins are of course not represented in the association between psychophysical scale values. Thus the elimination of these responses from the measures might represent an artificial inflation of agreement on the principle.

However, these missing data prove not to invalidate either the overall estimate of consensus on the principle or the assessment of the impact of demographic factors. The simplest summary indication that this interpretation is appropriate is provided by correlations between number of missing responses on psychophysical values (including both write-ins and simple blanks) and all other measures. For both the abstract and vignette crimes, the number of missing psychophysical values was uncorrelated with dependent measures indicative of fitting the model: i.e., with both betas and with both r 's. Thus for the majority of respondents, having missing data simply represented using lesser punishments than were provided in the task, but within the same overall model of fitting the punishment to the crime. Missing data at the individual level has no implication for the overall model. Further, the number of

missing responses also had only one significant demographic correlate (with sex on the vignette crimes), indicating that the conclusions about social status need no modification. In summary, the punishment scale provided appeared to be overly harsh from most respondents' point of view, as evidenced by their willingness to write in less harsh alternatives at least once. But what they appeared to want to do was to extend the game rather than change it.

Conclusions

To measure a norm is a complex task. In the present paper we first distinguished between norms as evidenced in behavior and norms as evidenced in the cognitions of society's members. In focusing on norms as cognitions held by individuals, we then distinguished between simple norms-- relating to specific acts or prohibitions-- and complex norms-- those relating sets of acts or objects to one another. Among these latter complex norms are included norms of reciprocity, exchange, and sanctioning: in short, norms of justice. We argued that social consensus on such norms can be conceptually broken down into consensus on the principle involved and consensus on the specific social facts or objects of evaluations. The conceptualization and the measurement of consensus on these norms must involve consideration of both potential levels of agreement.

To measure consensus on norms of justice involves a combination of within-individual and between-individual analyses. To assess whether individual members of the public hold to such a norm as a principle involves using each respondent as an experiment in model-fitting. One needs to know how the individual assessed the inputs and the outcomes to be exchanged or matched;

one needs to know how the individual matches them; and one needs to have a model of the appropriate, proportionate, or just matching against which to assess the individual's "model." Consensus on a norm of justice as a principle is to be found within individual members of the public. Consensus on the particulars of social stimuli, in contrast, can be assessed by more standard measures of agreement across individuals. Members of the public may agree on the principle, on the particulars, on both, or on neither. Disagreement may be either random or demographically linked in what is ordinarily thought of as a conflict situation.

The complex norm tested in the present data analysis was the notion of just deserts in criminal sanctioning, the idea that the punishment should fit the crime. We indicated that this norm has actually served as the untested underpinning of prior investigations of consensus on criminal justice, as in investigations of seriousness of crime. This norm can also be viewed as a special case of the notion of equity, a topic which has received much recent theoretical and empirical attention within social psychology. Despite extensive investigation, equity itself has never been tested in the appropriate manner described above. The present paper thus represents a first test of equity itself as well as of the specific substantive issue of whether the punishment fits the crime.

In order to test just deserts as a norm, psychophysically validated ratio scale measures of crime seriousness and punishment severity were used as the indicators of individuals' subjective evaluations of the inputs and outcomes to justice. The model fitted, derived from both psychophysics and equity theory, was one of proportional match between the seriousness of a crime and the severity of punishment assigned to it. Results indicated that members of

a block quota sample of the Boston SMSA did indeed fit the punishment to the crime, but according to a looser monotonic rather than proportional matching. Within-individual correlations between crime seriousness and punishment severity were substantial, however... Between-individual agreement on the seriousness of the crimes and the severity of the punishments employed was also substantial.

Two quite distinct forms of dissent were identified, however. A path model of the relation between aggregate consensus on stimuli and individual following of the principle revealed the existence of what we have called principled dissent: that when any disagreement exists about the evaluation of specific stimuli, the more an individual holds to the principle the more systematic that individual's deviation from the group average. In predicting individual relations of justice from aggregate relations, correlated errors in a path model are indications that the individual imposes greater cognitive structure than would be predicted from consensus in the sense of agreement with group averages. Thus consensus on principle implies principled dissent.

A second version of dissent, the more familiar demographic cleavage, was also identified on both adherence to the principle and agreement with group averages. The consistent predictors of not fitting the punishment to the crime were telling ones: income and race. Lower income or black respondents were less likely to exhibit the high correlations between crime seriousness and punishment severity which otherwise pervaded the data set. The absence of education effects on these same correlations makes it unlikely that the disagreements by income and race are effects of either sophistication at such tasks or knowledge of the relevant norms. Income and race also related to divergence / group averages in evaluations of the stimuli involved. Higher income respondents were more likely to agree with the average on

punishments, while black respondents were less likely to agree with the average on crime seriousness. Thus conflict-- of a social structurally meaningful variety-- was identified both on adherence to just deserts itself and on evaluation of the social stimuli making up the equation of the punishment and the crime.

By either the criterion of consensus on principle or that of consensus on social facts, just deserts appears to warrant being called a norm-- despite the evidences of conflict found. But to measure a norm is not to confirm a norm, and we are struck by the possible meaning of those evidences of conflict. It is still possible that disagreeing respondents did not understand the task at hand, but it seems at least as likely that they understood it all too well. To measure a single norm involves limitations in what is presented and in what can be found. The task presented was one which evoked, and essentially allowed, only a single frame of reference for justice and only a single basis of justice: desert.

Just deserts is not the only principle of justice. As Anatole France remarked, "The law, in its majestic equality, forbids the rich as well as the poor to sleep under bridges, to beg in the streets, and to steal bread." This tension, the injustice of justice, pervades the standards of a system of criminal law. Equity in meting out punishments according to desert can clash with prior inequity in meting out rewards according to desert. Or need. Or rights. The bulk of equity research, including our own, has focused on allocation of rewards or punishments according to a single principle, that of desert. In discussing the related area of exchange theory, Heath (1976) has characterized the research as being singlemindedly focused on desert as the sole basis of justice. Heath concludes that such a focus is not the only

potentially applicable rule, but that alternative and potentially competing rules can involve need or rights rather than desert. A focus on desert, Heath argues, is particularly likely to emerge in a liberal capitalist state. As Homans (1976) noted (in speaking of equity theorists' optimism rather than their capitalism), "How liberal and how American these assumptions are!" Market principles of resource distribution, as opposed to principles that evoke rights to goods or needs for goods, have long been at the core of our social arrangements. And market arrangements are a form of stabilized power differences (Giddons, 1973). Individual acts that redistributively circumvent these arrangements are regarded as crimes. To turn around the statement of demographic differences, we could say that highest agreement with fitting the punishment to the crime was shown by those who can be presumed to benefit most from the sanctioned distributive arrangements of work and contract, the high income or white respondent.

Thus competing principles of justice can readily be found. In the present study, only a single organizing principle was available; demographic disagreements were actually tendencies to fit the single model less well. This study, like all others in the literature, evoked only a single frame of reference for judging what is equitable. One can certainly imagine an alternative purified frame, such as that of need. The allocation of benefits to the needy could be evaluated in a research instrument modeled on the one employed in this study, for example, by describing cases from the New York Times annual Hundred Neediest (perhaps anchoring the scale with an unneedy plutocrat or two). It can also take very little to evoke a shift of frame within the respondent to an alternative principle of justice. This would presumably be accomplished by then describing one of the recently ranked neediest as the

perpetrator of a brutal mugging. One can even make the point of view really spin, by then describing the victim as a wealthy misanthrope who just laughingly bounced a large check to the Times charity.

Presumably, respondents would become quite angry if they were subjected to such antics, but their frustration would only be matched by that of the data analyst trying to make sense out of such nonsense. It would obviously be poor research design to haphazardly include descriptions whose various elements evoked different frames of reference with respect to justice. But it is also nonsense for the researcher who has demonstrated that respondents can and do use one frame of reference to conclude that it is the only one.

The present study investigated a single norm and evoked a single frame of reference from most respondents. We cannot know the reason for lower adherence to the principle of just deserts among low income and black respondents, although we can suspect that the answer has to do with alternative frames of reference for justice. Careful investigation of competing principles, as opposed to the helter-skelter frame shift exercise described above, could help to clarify such issues. Allocation of punishments could be tested for evidence of use of a principle like need as well as the principle of desert. Allocation of distributive rewards could also be examined as a problem in competing justice rules. If they are asked, respondents might prove to utilize quite sophisticated combinatorial solutions to the problem of justice in a multi-principled world; they might even judge inputs as a function of ability while they judge outputs as a function of need. What is most likely is that they will struggle with the various potential inputs and outputs in ways that reflect the complexity of multiple competing principles of justice. Competing justice norms, rather than disagreement about a single norm, may prove to be the crucial question for students of social consensus.

Footnotes

1. The statistical frequency or infrequency of a behavior is a guide to the prescription or proscription of certain ways of acting. Sanctioning, whether positive or negative, is a societal mechanism for enforcing appropriate behavior. And codification, the embodiment of ideals in written codes, is a strategy for long-term maintenance and transmission of the norm.
2. Among the variety of confirmatory studies are those by Akman, Normandeau, and Turner (1967), Normandeau (1966), and Velez-Diaz and Megargee (1970).
3. Three other studies have addressed both crime seriousness and punishment severity. Each appears to assume the norm in ways similar to Sellin and Wolfgang (1964) or Rossi et al. (1974). The earliest study, by Rose and Prell (1955), had respondents do Thurstone scales for both seriousness of crimes and severity of punishments. The major point of the article was that rated seriousness of the 13 minor felonies they used as stimuli did not correspond to ranges of actual punishments, which they took as evidence of clash between law and normative structure. The authors also had respondents actually assign punishments to the crimes, but focused only on demographic differences in rates and effect of manipulated offender characteristics in analyzing punishment assignments. Gibbons (1969) was also interested in discrepancies between public sentiment and legal practices. His indicator of the perceived seriousness of offenses was the punishments which respondents assigned to them. Finally, White (1975) investigated effects of manipulated offender and victim status on severity of punishment. He found negligible effects of these variables, but a substantial effect of crime seriousness on punishment. In his study, seriousness was operationalized by assigning Sellin-Wolfgang ratings to the stimuli;

punishment severity was the ordinal 12-point scale employed by Gibbons. Thus White's study differs from the present one in his substantive focus and in his imposition of scale scores rather than obtaining subjective ratings from the actual respondents.

4. Brickman (1977) has recently presented an alternative conception of how criminal justice might be organized around victim restitution rather than offender punishment. The restitution version of criminal justice is more closely linked to standard equity theory formulations of justice, as well as to the structure of Anglo-American civil law. We should stress that the description of criminal justice "equity" in the present paper is based on the way the system appears to be currently organized, and does not reflect our views of how it should be organized.
5. In assessing the distributive justice of pay outcomes, Jasso and Rossi (1977) have recently produced a methodological and conceptual advance by exploring the effect of various real-world characteristics (e.g., sex and marital status) on whether pre-specified earnings are perceived as fair by adults. Thus they have begun to assess the bases of the underlying distribution rule itself. In addition, they explore consensus in terms of demographic differences on these judgments. However, their measurement technique does not allow them to make use of their within-respondent design in the manner suggested below; to do so they would need to obtain respondents' own subjective judgments of the inputs and outcomes per se and respondents' own assignments of outcomes for inputs.
6. Early investigations of this type tended to validate the scales obtained against other category scales for the same dimension or against other relevant variables. Sellin and Wolfgang's (1964) use of the correlation of their scale measures with punishments in Pennsylvania is an example of such

validation procedures. More recent procedures for validating psycho-physical scales for social opinion rest on a procedure used for physical stimuli known as cross-modality matching. With physical stimuli, cross-modality matching consists of having subjects adjust one scalable stimulus to match another (e.g., adjusting the perceived brightness of a light to match the weight of an object); the function derived reflects the ratio of the two individual modalities. For opinions about a social stimulus, indirect cross-modality matching can be used as a validation technique. Subjects indicate the intensity of their judgments of a social stimulus by adjusting responses to one physical modality, matching the strength of response to the intensity of opinion; second, they repeat the procedure using a different physical modality; finally, an indirect cross-modality match is achieved by plotting the responses made to the same social stimulus on the two modalities against one another. If the social stimuli indeed constitute a single ratio scalable dimension, the resulting match produces a power function whose exponent is approximately the ratio of the characteristic exponents of the two modalities (see, e.g., Dawson and Brinker, 1971).

7. Each respondent is allowed to set a standard or "modulus" with the first response given by assigning a number to the first stimulus in ME judgments and drawing a line for the first stimulus in LP judgments. These are supposed to represent the seriousness/severity of that stimulus. All other stimuli are judged relative to the first; respondents give numbers or draw lines to represent the proportional relations of the stimuli. For example, if they feel a crime is twice as serious as the first they should assign a number twice as big or draw a line twice as long. The data are then analyzed as described in the text.

8. Blocks were drawn in proportion to representation in the 1970 census and an additional eight blocks were selected in which only black respondents were eligible, based on 1970 census blocks with 20% or higher black populations. This strategy was chosen to ensure adequate black representation in the overall sample. Interviewers were assigned randomly to blocks, with quotas imposed for sex and age of respondents. A total of 391 interviews were obtained of an original goal of 400 interviews.
9. These technical corrections are for what is known as regression error. Cross (1974) and Lodge *et al.* (1974) have developed procedures for obtaining unbiased estimates of the regression errors on the physical variables used in cross-modal matching. This methodological advance frees the researcher from the assumption of no error in the regression of one physical variable against the other, provides a criterion for accepting or rejecting the results of cross-modal matches, and provides a means for deriving unbiased scales of opinion from these sensory matching functions. The present paper did not employ regression corrections, because these are made on group-level rather than individual-level data. Previous analyses have shown regression bias to be minimal in the present data set (Hamilton and Rotkin, 1978).
10. A consistent non-zero beta of whatever magnitude would be indicative of a law-like fitting of seriousness to severity. A beta equal to unity would describe a fit that corresponds to the usual notion of proportionality. Under such a fit, a doubling of the seriousness input would call for a doubling of the severity output. A beta equal to "two" would indicate a fit where a doubling of the seriousness input would call for a quadrupling of the severity output. A beta equal to 1.73 would mean that a doubling of seriousness was matched with an increase in severity equal to $2^{1.73}$. Such

power laws with non-unit exponents are common in the subjective scaling of physical inputs that is the central concern of psychophysical scalars. For example, doubling the loudness of a sound source results in a subjective increase in intensity of $2^{.67}$.

11. The n for these analyses is reduced from 314 to 297 because we include only those respondents who have both seriousness and punishment scores for both abstract and vignette crimes. This includes respondents who have some missing data but for whom regression lines can still be computed.
12. We thank Robert Zajonc for removing the Durkheimian wool from our eyes and forcing us to acknowledge the potential spuriousness of seriousness judgments. In addition to the evidence provided against this in the text, evidence across prior studies is also supportive of the view that seriousness is a judgment distinct from a consideration of typical punishments. In Hamilton and Rotkin (1976), the ME and LP seriousness judgments followed the punishment judgments; in Hamilton and Rotkin (1978), they preceded punishment judgments. The similarity of scale range and rank ordering between the two data sets suggests that seriousness judgments per se are unaffected by considering punishments per se.
13. This notion of creating quasi-independent replicates was suggested by Kenneth Guire of the Statistical Research Laboratory, University of Michigan. The problem was especially difficult in these data because of the way in which they were prepared. Since we subtracted out the mean of each individual's logged responses, it is impossible for individual differences to be estimated by an additive model.
14. One alternative explanation of deviations, coding error, can be dismissed. Examination of a subset of highly deviant cases revealed no errors in coding, such as a reversal of one scale or another, to provide a comfortable

explanation of the sometimes zero and occasionally negative relationships found. Aside from sheer idiosyncrasy (which individual questionnaires exhibited in numerous creative ways), differences would therefore appear likely to be demographically caused.

15. Rossi, Waite, Bose and Berk (1974) performed a similar analysis on squared correlations of individuals with aggregate seriousness ratings. The rationale for squaring the correlation coefficients was not offered but it does reduce the skew in the marginal distribution and results in a metric with which sociologists are quite at home. However, the statistical properties of the square transformation are not as desirable as the Fisher's z transformation, which yields a theoretically normal distribution, and squaring is obviously inappropriate when both positive and negative correlations are found, a problem which exists for our data but is trivial in Rossi *et al.*'s. For comparative purposes, we performed an analysis like that of Table 5 using squared correlation coefficients. The pattern of signs is the same and all of the statistically significant correlations of Table 5 remain so. In addition, in the vignette portion of the table, three additional significant correlations were observed, race with r_{pp} ($r = -.12$) and education with r_{aS} ($r = .16$) and r_{pP} ($r = .12$). This analysis offers mixed replicative support for the Rossi *et al.* finding of a substantial education effect on the squared correlation of individual and aggregate seriousness, with a significant effect found only for the vignette data and not for the abstract label ratings which more closely resemble the Rossi *et al.* stimuli. The statistically more appropriate Fisher's z transformation yields no support. One can only speculate why our seemingly more complex measurement strategy failed to reveal education-linked differences that Rossi *et al.* observed.

16. Although alpha would appear to be a stable indicator of punitive tendencies, neither alpha nor our measure of average punishment assigned showed stable relationships to any demographic variable other than race. Race showed a consistent (although not always significant) pattern for blacks to assign less punishment. Thus we conclude that the chief demographic differences in this data set are with respect to fitting punishment to crime or agreeing with the group averages rather than punitiveness per se.

References

- Adams, J.S. 1965. "Inequities in Social Exchange." Pp. 267-299 in Advances in Experimental Social Psychology (Vol. 2), edited by L. Berkowitz. New York: Academic Press.
- Adams, J.S. and S. Freedman. 1976. "Equity Theory Revisited: Comments and Annotated Bibliography." In Advances in Experimental Social Psychology (Vol. 9), edited by L. Berkowitz and E. Walster. New York: Academic Press.
- Akman, D.D., A. Normandeau, and S. Turner. 1967. "The Measurement of Delinquency in Canada." Journal of Criminal Law, Criminology and Police Science 58: 330-7.
- Anderson, T.W. 1958. An Introduction to Multivariate Statistical Analysis. New York: Wiley.
- Austin, W., E. Walster, and H.K. Utne. 1976. "Equity and the Law: The Effect of a Harmdoer's 'Suffering in the Act' on Liking and Assigned Punishment." In Advances in Experimental Social Psychology (Vol. 9), edited by L. Berkowitz and E. Walster. New York: Academic Press.
- Berkowitz, L. and E. Walster (Eds.). 1976. Advances in Experimental Social Psychology (Vol. 9). New York: Academic Press.
- Biggs, J. 1955. The Guilty Mind: Psychiatry and the Law of Homicide. Baltimore, Md.: Johns Hopkins Press.
- Brickman, P. 1977. "Crime and Punishment in Sports and Society." Journal of Social Issues 33: 140-64.
- Cliff, N. 1973. Annual Review of Psychology 24: 473-506.
- Cross, D.V. 1974. "Some Technical Notes on Psychophysical Scaling." In Sensation and Measurement: Papers in Honor of S.S. Stevens, edited by H. Moskowitz, B. Scharf, and J.C. Stevens. Dordrecht, Netherlands: Reidel.
- Dawson, W.E. and R.P. Brinker. 1971. "Validation of Ratio Scales of Opinion by Multi-modality Matching." Perception and Psychophysics 4: 413-7.
- Duncan, O.D. 1975. Introduction to Structural Equation Models. New York: Academic Press.
- Gibbons, D.C. 1969. "Crime and Punishment: A Study in Social Attitudes." Social Forces 47: 391-7.
- Giddons, A. 1973. The Class Structure of the Advanced Societies. New York: Harper and Row.
- Gouldner, A.W. 1960. "The Norm of Reciprocity: A Preliminary Statement." American Sociological Review 25: 161-78.
- Hamblin, R.L. 1971. "Mathematical Experimentation and Sociological Theory: A Critical Analysis." Sociometry 34: 423-52.
- Hamilton, V.L. and L. Rotkin. 1976. "Interpreting the Eighth Amendment: Perceived Seriousness of Crime and Severity of Punishment." In Capital Punishment in the United States, edited by H.A. Bedau and C. Pierce. New York: AMS Press.
- _____. 1978. "The Capital Punishment Debate: Public Perceptions of Crime and Punishment." Accepted for publication, Journal of Applied Social Psychology.
- Harris, R. 1976. "Handling Negative Inputs: On the Plausible Equity Formulae." Journal of Experimental Social Psychology 12: 194-209.
- Hart, H.L.A. 1968. Punishment and Responsibility. New York: Oxford University Press.
- Heath, A. 1976. Rational Choice and Social Exchange: A Critique of Exchange Theory. Cambridge, Eng.: Cambridge University Press, 1976.

- Hodge, R.W., P.M. Siegel, and i.H. Rossi. 1964. "Occupational Prestige in the United States, 1925-1963." American Journal of Sociology 70: 286-302.
- Homans, G. 1976. "Commentary." In Advances in Experimental Social Psychology (Vol. 9), edited by L. Berkowitz and E. Walster. New York: Academic Press.
- Jasso, G. and P.H. Rossi. 1977. "Distributive Justice and Earned Income." American Sociological Review 42: 639-51.
- Kalven, H. and H. Zeisel. 1966. The American Jury. Chicago: University of Chicago Press.
- Kuennapas, T. and I. Wilkstroem. 1963. "Measurement of Occupational Preferences: A Comparison of Scaling Methods." Perceptual and Motor Skills 17: 611-24.
- Lodge, M., B. Tursky, J. Tannenhaus, and D. Cross. 1974. "The Development and Validation of Political Attitude Scales: A Psychophysical Approach." Paper delivered to National Science Foundation Conference on Design and Measurement Standards for Research in Political Science. Delevin, Wisconsin, May 13-15.
- Lodge, M., J. Tannenhaus, D. Cross, B. Tursky, M.A. Foley, and H. Foley. 1976. "The Calibration and Cross-modal Validation of Ratio Scales of Political Opinion in Survey Research." Social Science Research 5: 325-47.
- McNemar, Q. 1969. Psychological Statistics (4th ed.). New York: Wiley.
- Nadel, S.E. 1957. The Theory of Social Structure. New York: Free Press of Glencoe.
- Normandeau, A. 1966. "The Measurement of Delinquency in Montreal." Journal of Criminal Law, Criminology and Police Science 57: 172-7.
- Rainwater, L. 1971. "The Measurement of Social Status." Department of Sociology, Harvard University.
- Robinson, W.S. 1950. "Ecological Correlations and the Behavior of Individuals." American Sociological Review 15: 351-7.
- Rose, A.M. and A.E. Prell. 1955. "Does the Punishment Fit the Crime? A Study in Social Valuation." American Journal of Sociology 61: 247-59.
- Rossi, P., E. Waite, C.E. Bose, and R.E. Berk. 1974. "The Seriousness of Crimes: Normative Structure and Individual Differences." American Sociological Review 39: 224-37.
- Sellin, J.T. and M.E. Wolfgang. 1964. The Measurement of Delinquency. New York: Wiley.
- Shinn, A.M., Jr. 1969. "An Application of Psychophysical Scaling Techniques to Measurement of National Power." Journal of Politics 31: 932-51.
- Simon, R.J. 1967. The Jury and the Defense of Insanity. Boston: Little, Brown.
- Stevens, S.S. 1975. Psychophysics. New York: Wiley.
- Velez-Diaz, A. and E.I. Megargee. 1970. "An Investigation of Differences in Value Judgments between Youthful Offenders and Non-offenders in Puerto Rico." Journal of Criminal Law, Criminology and Police Science 61: 549-53.
- von Hirsch, A. 1976. Doing Justice: The Choice of Punishments. New York: Hill and Wang.
- Walster, E., E. Berscheid, and G.W. Walster. 1973. "New Directions in Equity Research." Journal of Personality and Social Psychology 25: 151-76.
- Walster, E., G.W. Walster, and E. Berscheid. 1978. Equity: Theory and Research. Boston: Allyn and Bacon.
- White, G.F. 1975. "Public Responses to Hypothetical Crimes: Effect of Offender and Victim Status and Seriousness of the Offense on Punitive Reactions." Social Forces 53: 411-9.
- Williams, R. 1968. "The Concept of Norms." In International Encyclopedia of the Social Sciences, edited by D. Sills. New York: Macmillan.

Table 1

Crime and punishment stimuli for psychophysical scaling tasks.^a

Crime list: Abstract Crimes	Crime List: Vignette Crimes	Punishment List
Taking \$50	Taking \$50	2 yrs. in prison
Stealing and abandoning an auto	Assault without weapon	3 yrs. in prison
Assault without a weapon	Assault with weapon	5 yrs. in prison
Mugging	Mugging	7 yrs. in prison
Politician accepting bribes	Armed robbery	15 yrs. in prison
Assault with a weapon	Forcible rape (1)	20 yrs. in prison
Spying for a foreign government	Manslaughter	25 yrs. in prison
Manslaughter	Forcible rape (2)	Life, parole after 25 yrs.
Armed robbery	Impulsive killing	Life, no parole
Hijacking a plane	Kidnapping	Death penalty
Forcible rape	Armed robbery/murder	
Impulsive killing	Planned killing (1)	
Kidnapping	Planned killing (2)	
Sale of heroin resulting in a death	Rape/murder	
Armed robbery in which victim is killed		
Planned killing		
Forcible rape in which victim is killed		

^aEach list is arranged in order of average seriousness on severity as determined by this sample. All lists were presented to respondents in random order.

Table 2
The individual-level model of "just deserts":
Correlations within respondents of crime
seriousness and punishment severity.

	Abstract crimes	Vignette crimes	Correlation between abstract and vignette
Average \bar{r}^a	.67	.79	.33
Median	.74	.85	
Minimum	-.67	-.62	(.297)
Maximum	.99	.99	
Standard deviation	.27	.23	
Average beta	.94	1.19	.25
Median	.86	1.10	
Minimum	-.66	-.61	(.297)
Maximum	7.22	4.23	
Standard deviation	.73	.72	
Average alpha	-.158	-.136	.71
Median	-.09	-.08	
Minimum	-3.40	-4.35	(.297)
Maximum	1.30	1.39	
Standard deviation	.49	.53	

N = 297

^a \bar{r} 's were normalized by Fisher Z -transform before being correlated with one another.

Table 5

Zero-order missing data correlations of relevant demographic variables with measures of fitting the punishment to the crime and of agreement with the average.^a

Where r_{sp} = correlation of individual's crime seriousness with severity
 r_{sS} = correlation of individual's crime seriousness with group average seriousness
 r_{pP} = correlation of individual's punishment severity with group average severity.
 partial r = individual's partial correlation estimated from path model 1(b), Figure 1
 alpha = intercept of individual's regression line
 meanpun = average punishment severity assigned by individual

Abstract Crimes

	Demographics ^b			Sex
	Race	Income	Education	
r_{sp}	-.15 ^d (299)	+.22 ^d (281)	+.11 (293)	-.07 (300)
r_{sS}	-.15 ^d (299)	+.01 (281)	+.08 (293)	+.07 (300)
r_{pP}	-.13 ^c (299)	+.19 ^d (281)	+.05 (293)	-.04 (300)
partial r	-.03 (298)	+.16 ^d (280)	+.07 (292)	-.17 ^d (299)
alpha	-.00 (299)	-.00 (281)	-.10 (293)	+.07 (300)
meanpun	-.12 ^c (299)	+.04 (281)	-.11 (293)	+.06 (300)

Vignette Crimes

r_{sp}	-.07 (301)	+.09 (282)	+.03 (295)	-.125 ^c (302)
r_{sS}	-.14 ^c (302)	+.07 (283)	+.08 (296)	+.02 (303)
r_{pP}	-.08 (301)	+.16 ^d (282)	+.08 (295)	+.02 (302)
partial r	-.02 (295)	+.02 (276)	-.00 (289)	-.185 ^d (296)
alpha	-.11 ^c (303)	-.01 (284)	-.09 (297)	+.16 ^d (304)
meanpun	-.11 ^c (303)	+.04 (284)	-.09 (297)	+.05 (304)

Table 5 (Continued)

^a All within-individual correlations were normalized by Fisher's z-transform before being correlated with the demographics. Nonsignificant correlations are indicated in smaller print to facilitate presentation. N's for each missing data correlation are presented below it in parentheses.

^b Race and sex are coded as dummy variables, with black = 1 and female = 1. Education is coded in six categories from low to high and income in 15 categories from low to high.

^c $p < .05$

^d $p < .01$

Table 3

Relation between crime seriousness and punishment severity across respondents, aggregated and disaggregated

S = group average seriousness for each crime
 P = group average punishment for each crime
 s = individual respondent's seriousness for each crime
 p = individual respondent's punishment for each crime
 D = dummies for which crime stimulus is being rated

	Abstract Crimes	Vignette Crimes
1. Regression of p on s:	R = .56	R = .66
	R ² = .31	R ² = .43
	coefficient = .73	coefficient = .94
2. Regression of p on s plus D:	R = .65	R = .70
	R ² = .43	R ² = .49
	partial r for s = .27	partial r for s = .38
	coefficient for s = .37	coefficient for s = .61
3. Regression of p on s plus S:	R = .64	R = .69
	R ² = .40	R ² = .48
	partial r for s = .26	partial r for s = .39
	coefficient for s = .37	coefficient for s = .62

N for all analyses = 4,595 responses for abstract crimes; 3,742 responses for vignette crimes.

Table 4

Summary of individual-level correlations of deviations from aggregate measures (partial correlation of crime seriousness with punishment severity controlling for sample average seriousness and severity).

	Abstract Crimes ^a	Vignette Crimes ^a
Mean	.35	.54
Median	.40	.62
Minimum	-.68	-.75
Maximum	.95	.99
Standard Deviation	.34	.33
% Greater than Zero	82%	93%

^aN = 297

Table 6

Deviation from the model of "fitting the punishment to the crime" by race and by income.^a

A. Race	White	Black	
"Bad" fit	17%	33%	18%
	45	9	54
"Good" fit	83%	67%	82%
	225	18	243
	100%	100%	100%
	N = 297		

$\chi^2 = 4.58$ $p = .03$

B. Family Income	Low (0-\$7,999)	Middle (\$8,000-\$19,999)	High (\$20,000 up)	
"Bad" fit	31%	18%	11%	19%
	22	18	12	52
"Good" fit	69%	82%	89%	81%
	49	80	98	227
	100%	100%	100%	100%
	N = 279			

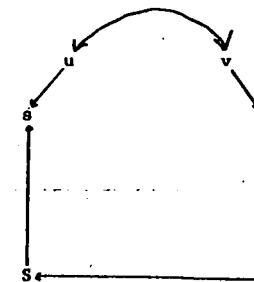
$\chi^2 = 11.47$ $p = .003$

^a"Bad" fit = correlation between rated crime seriousness and assigned punishment severity < .5 for either abstract or vignette stimuli. "Good" fit = both correlations \geq .5.

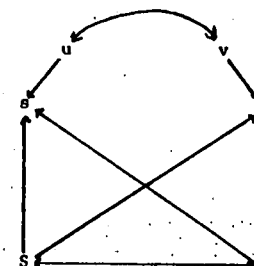
Figure 1

Models of the determination of an individual's seriousness-severity correlations by sample average ratings.

Model 1(a)



Model 1(b)



where S = the sample average seriousness ratings
P = the sample average severity ratings
s = the individual's seriousness ratings
p = the individual's severity ratings
u, v are the deviations

