

Purposes of Punishment: Effects of Utilities of Criminal Sanctions on Perceived Appropriateness

Robert M. McFatter
University of Southwestern Louisiana

Five commonly discussed purposes of punishment (incapacitation, retribution or just deserts, rehabilitation, general deterrence, and special deterrence) were examined in two experiments (using college students and six district judges). The experiments were designed to study how the perceived usefulness or utility of penalties in serving these purposes changes with seriousness and type of crime, and how these perceived utilities affect judgments about the appropriateness of penalties for various crimes. Subjects rated the usefulness of 30 possible penalties (including fines, probation terms, and imprisonment terms) in serving these purposes for four crimes. The implications of the obtained utility functions in understanding conflicting sanctioning norms are discussed. Importance weights derived from the data indicated just deserts to be the most influential for all crimes among the judges and for all crimes except murder among the students. The results suggest that it is important to distinguish between what decision makers say the most important considerations ought to be and how their decisions actually appear to depend on these factors.

A sentence imposed on a criminal offender could be intended to accomplish any or all of a number of objectives (see, e.g., Ezorsky, 1972; Hart, 1968; Vidmar & Miller, 1980). This article investigates (a) perceptions of the usefulness of various penalties in serving five commonly discussed purposes of sanctions, and (b) how these perceptions affect judgments about the punishment appropriate for a criminal offender.

The five purposes examined here are (a) *incapacitation*: the attempt to physically prevent the offender from committing further crime by restraining him in some manner (usually imprisonment); (b) *retribution/just deserts*: the attempt to impose a penalty on the offender strictly in accordance with what he deserves because of the crime he has committed; (c) *rehabilitation*: the attempt to "treat" the offender or change him through corrective measures so that he will not com-

mit further crime and can be safely returned to society; (d) *general deterrence*: The attempt to impose a penalty on the offender such that other potential offenders in the general public will not commit crime through fear of punishment; and (e) *individual or special deterrence*: the attempt to impose a penalty on the offender such that he will not commit further crime through fear of punishment.

These five purposes are among the most commonly cited justifications for imposing specific sanctions on offenders and would seem to be important considerations in understanding how people, including judges, legislators, and the general public, make judgments about what punishment a given offender ought to receive. Indeed, many people no doubt see the sentencing task as primarily one of finding a penalty that will accomplish a wise blending of these purposes.

Several studies have examined social consensus on the norm of just deserts in criminal sanctioning (Hamilton & Rytina, 1980; Rossi, Waite, Bose, & Berk, 1974; Sellin & Wolfgang, 1964). These studies demonstrate that there is a good deal of agreement both on the principle that penalties ought to be proportional to the seriousness of the crime and on perceptions of the relative se-

Portions of this paper were presented at the 27th Annual Meeting of the Southeastern Psychological Association in Atlanta, Georgia, March 1981. The author wishes to thank Kathryn Ashworth for help in collecting the data. Helpful comments were provided by Harry F. Gollob and Richard M. McWhirter.

Requests for reprints should be sent to Robert M. McFatter, Psychology Department, Box 4-3131, University of Southwestern Louisiana, Lafayette, Louisiana 70504.

riousness of many crimes. As Hamilton & Rytina (1980) point out, however, "the results do not tell us about the relative strength of just deserts as a norm or about people's willingness to use other alternative principles of justice" (p. 1141). Thus, an important issue in any theory of social norms about criminal sanctions is the question of the relative strength of the various competing goals in affecting perceptions of the appropriateness of penalties.

There have been a few studies looking at the relative importance subjects *claim* to place on the various purposes and how this is related to recommended sanctions (Hogarth, 1971; McFatter, 1978; Vidmar, 1974), but there has been no effort to systematically examine common perceptions of how well various penalties fulfill the five purposes, much less how these perceptions actually affect judgments about the appropriateness of possible penalties.

The approach adopted in the present investigation is to assume that any possible sanction that might be imposed on an offender may be thought of as fulfilling each of the five purposes about to varying degrees. Thus, it is possible to speak of the usefulness or utility of a sentence of 2 years imprisonment for incapacitation, and similarly its utility for retribution, rehabilitation, and so on. For example, a \$500 fine might have no utility for the incapacitation of an offender, but might have substantial utility as a general deterrent for certain crimes. If measures of the perceived utility of a variety of penalties for each of the purposes above can be obtained, it becomes possible to examine not only how these perceived utilities vary from crime to crime, but also how they are combined to produce judgments about the overall appropriateness of any possible penalty. The two experiments described below were designed to obtain such measures.

A number of questions arise when the problem is formulated in this way. For instance, is the rehabilitative utility of a 3-year probation term perceived to be the same for a car theft offender and a rape offender? In general, how do perceptions of the utility of penalties for the various purposes change from crime to crime? Do perceptions of the appropriateness of various penalties for a

crime appear to rely on the penalties' utility for a number of purposes or only one or two?

Judgment Model

A common and useful way of analyzing how information is used by a decision maker involves the construction of a statistical model (usually linear) that describes how the individual attributes of a complex stimulus are weighted and combined to produce a global judgment about the stimulus. The stimuli investigated in the experiments below comprise a wide set of possible penalties that might be imposed on a criminal offender. The attributes presumed to influence a penalty's perceived appropriateness are its perceived utilities for each of the five purposes.

It will be assumed that the perceived overall appropriateness of the *i*th penalty for a given offender can be described by the following equation:

$$A_i = \sum_{j=1}^k w_j U_{ij}, \quad (1)$$

where A_i = judged appropriateness of the *i*th penalty, the U_{ij} s are perceived utilities of that penalty for each of the *k* purposes, and the w_j s are relative importance weights attached to each purpose. Linear models like Equation 1 have been used extensively in a wide variety of judgment situations (Hammond, McClelland, & Mumpower, 1980; Hammond, Stewart, Brehmer, & Steinmann, 1975; Slovic & Lichtenstein, 1971) to describe the policies and values of the decision makers involved.

The form of Equation 1 implies certain assumptions about how utilities are combined. For instance it is a linear compensatory model, which implies that a penalty's usefulness for one purpose can compensate for its lack of usefulness for another purpose. Other model forms will incorporate somewhat different assumptions about how utilities affect appropriateness. A model of the form

$$A_i = w_0 \prod_{j=1}^k U_{ij}^{w_j}$$

is sometimes called a conjunctive model (Einhorn, 1970) because it implies that a penalty must have at least minimal usefulness for all purposes to be seen as highly appropriate.

It is possible to derive other model forms to incorporate different assumptions about how utilities are combined. For example, it might be the case that it is only when penalties are reasonably fair (i.e., useful for just deserts) that other purposes become important considerations. Something like this appears to be what Morris (1974) had in mind in his prescriptive recommendations for sentencing decisions. It can be shown that a model of the form

$$A_i = U_{i,j} \sum_{j=1}^k w_j U_{ij}, \quad (2)$$

where U_j is utility for just deserts, has this property and thus might be a plausible candidate for a model to describe appropriateness judgments.

Preliminary examination of data from the experiments described below indicates that the patterns of weights estimated for the conjunctive model are similar to those of Equation 1. Interpretation of the weights in the model of Equation 2 is somewhat problematic, and the very structure of the model implies a priori that utility for just deserts is the most important influence on appropriateness. In addition, the fits of the nonlinear models to the obtained data are typically not as good as those of the linear compensatory model of Equation 1. Consequently, Equation 1 will be assumed to be an adequate description of the judgment process throughout this article.

There are several issues that can be examined by obtaining estimates of the importance weights in Equation 1. First, which of the purposes appear to have the most influence in determining the overall judged appropriateness of penalties, and does the relative influence of the purposes change from crime to crime? For example, it might be seen as more important to incapacitate rapists than to give them what they strictly deserve, whereas just the opposite might hold for car theft offenders.

Second, it is of interest to examine the

degree of correspondence between how subjects respond when directly asked to indicate how important each of the five purposes should be in determining an appropriate sentence and the weights estimated for Equation 1, representing in some sense the weights that subjects actually appear to use when deciding on the appropriateness of penalties. There is considerable evidence that both expert and nonexpert judges in a variety of different judgment tasks have little insight into which pieces of information most strongly affect their judgments (Nisbett & Wilson, 1977; Slovic & Lichtenstein, 1971).

Slovic & Lichtenstein (1971), in reviewing the judgment literature, found that "judges strongly overestimate the importance they place on minor cues . . . and they underestimate their reliance on a few major variables" (p. 684). They argued (as have others) that one reason judges may have little insight into which factors most strongly influence their judgments is that for highly practiced tasks judgment becomes almost automatic, requiring very little conscious attention to the processes involved. In the first experiment below, college student subjects were not expected to be highly experienced at picking appropriate penalties for criminal offenders. Nonetheless, it may be argued that the equity principle—that people should get what they deserve, or that their outcomes should be proportional to their inputs—is such a pervasive principle in social cognition that it becomes almost unconscious and people might well systematically underestimate the degree to which they use it in judging the appropriateness of penalties.

Experiment 1

Method

Subjects. A total of 86 undergraduate college students enrolled in introductory psychology participated in the experiment. Subjects were randomly assigned to make judgments about one of four crimes. Six subjects returned incomplete questionnaires (pages skipped, etc.) and were dropped from further analysis. Four subjects' questionnaires were unusable because they obviously misunderstood the instructions (e.g., assigning numbers outside the 0–10 scale in the directions). Four additional subjects, randomly selected, were dropped to provide equal numbers of subjects in each crime group. The

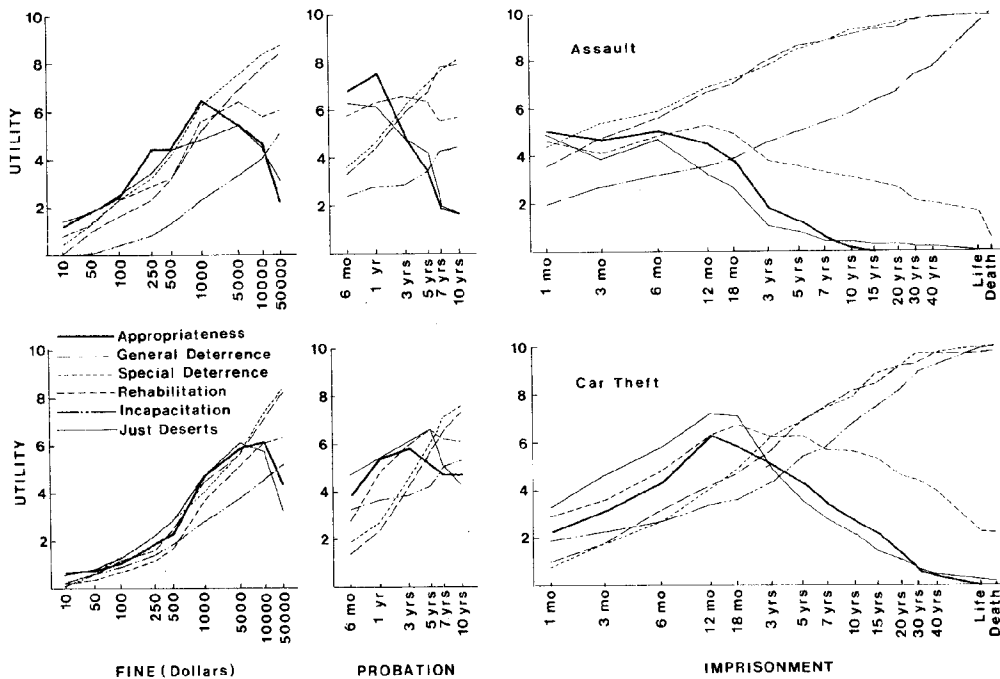


Figure 1. Aggregate utility functions for assault and car theft from college students' ratings. Penalties are scaled on abscissa using values from Sebba (1978) in all figures.

analysis was based on a total of 72 subjects, 18 in each crime group.

Crimes. Four short crime descriptions were used as stimuli for subjects' judgments. The four descriptions are given below along with the title that will be used to refer to that crime throughout this article.

(assault): The offender became involved in an argument with the victim after a minor traffic accident in which the victim was at fault. Witnesses said the offender became angry, physically attacked the victim, and beat him to the point that he required a short hospitalization. The offender is 25 yrs. old and has no previous criminal record.

(car theft): The offender stole a late-model car from a parking lot and totaled it in a police chase several hours later. The offender is 25 yrs. old and has no previous criminal record.

(rape): The offender followed a young woman home from the supermarket, talked his way into her apartment by pretending to be a maintenance man, and raped her at knife point. The offender is 25 yrs. old and has no previous criminal record.

(murder): The offender became involved in an argument with his wife in a bar. When a bystander tried to intervene, the offender pulled a knife and stabbed him to death. The offender is 25 yrs. old and has no previous criminal record.

Procedure. Subjects were run in groups of 20-30. They were first asked to fill out the Just World Scale

(Rubin & Peplau, 1973), after which questionnaire booklets were distributed for them to fill out. Each booklet contained one of the crime descriptions and instructions for the judgment task. All booklets were identical except for the crime descriptions, and the booklets were randomized before distribution so that subjects were randomly assigned to crime groups.

To facilitate comprehension of the nature of the judgments called for, the experimenter carefully read aloud all instructions to subjects except the crime descriptions. The instructions thoroughly explained each of the five purposes and asked subjects to rate the usefulness of each of 30 possible penalties in accomplishing each of the five purposes for the offender in the crime description. Although there have been a variety of methods proposed to elicit utility functions (Huber, 1974; Keeney & Raiffa, 1976; Wascoe, Note 1), many of these methods are cumbersome, time consuming, and irritating to the decision makers involved. There is serious question whether any potential formal benefits gained from their use over simpler methods are worth the cost and risk of "confused responses to complex questions" (Slovic, Fischhoff, & Lichtenstein, 1977, p. 23). For this reason and the large number of judgments required, the method adopted in the present investigation was to have subjects simply make ratings of the usefulness of penalties in accomplishing each of the purposes.

Ratings were made on 11-point (0-10) scales with the following anchors for the five purposes: (a) incapacitation (does not incapacitate offender at all/totally incapacitates offender); (b) just deserts (absolutely not

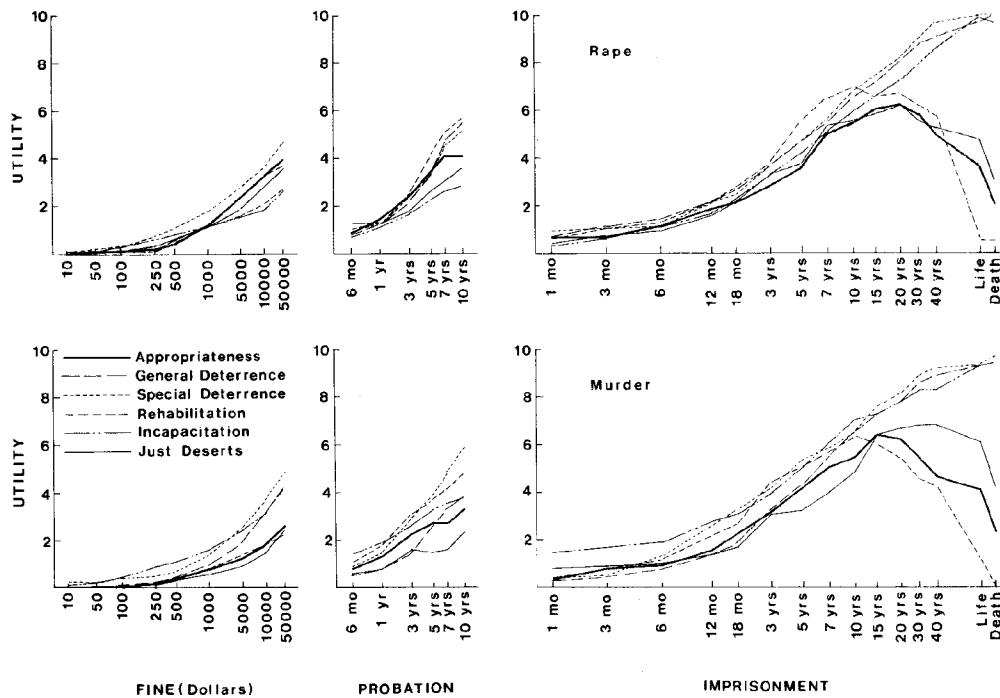


Figure 2. Aggregate utility functions for rape and murder from college students' ratings.

what offender deserves/exactly what offender deserves; (c) rehabilitation (will not rehabilitate offender at all/certain to rehabilitate offender; (d) general deterrence (will not deter offenders/will strongly deter offenders; (e) special deterrence (will not deter offender/will strongly deter offender).

Subjects were next asked to rate each of the 30 penalties with respect to how appropriate each would be, all things considered, as a sentence imposed on the offender in the crime description. Again, an 11-point (0-10) scale was used with anchors of totally inappropriate and highly appropriate.

Following this, subjects made judgments about a number of other aspects of the crime, offender, and victim such as seriousness of the crime, dangerousness of the offender, and attributions of blame. These judgments are not the focus of the present article and will not be discussed further.

Finally, subjects rated the degree of importance they personally felt should be attached to each of the five purposes in determining an appropriate sentence on the following 5-point scale: 0 = no importance, 1 = little importance, 2 = some importance, 3 = very important, 4 = most importance.

Penalties. It was important to have subjects rate a wide range of possible penalties that varied substantially in perceived utility for the various purposes and appropriateness to get an adequate idea of the effects of the perceived utilities over the entire range of possible penalties. A set of 30 possible penalties including 9 fines, 6 probation terms, 14 imprisonment terms, and the

death penalty was used in the present experiment. These penalties were a subset of the penalties examined by Sebba (1978) who was interested in scaling the relative severity of criminal sanctions. They were selected to cover as much as possible the full range of possible penalties that might be assigned by a sentencing judge. The 30 penalties are shown in Figures 1 and 2 where they are placed on the horizontal axis using scale values obtained by Sebba (1978).

Results and Discussion

Aggregate utility functions. Aggregate utility functions were constructed for each of the five purposes for each crime by averaging the ratings made by all subjects in each crime group for each purpose. Figures 1 and 2 show these functions for each crime along with the average overall appropriateness function.

It is clear from the figures that, for any given penalty type, the utility functions for incapacitation, general deterrence, and special deterrence are monotonically increasing functions of penalty severity, and this holds for all crimes. On the other hand, utility functions for just deserts and rehabilitation

appear to increase to some optimum penalty and then decrease. Further, the point of optimum utility gradually increases with increasing crime seriousness. (The average seriousness ratings for the crimes were: assault = 4.83, car theft = 5.06, rape = 8.11, murder = 8.67; $F[3, 68] = 28.04, p < .001$.) The overall appropriateness function tends to follow the just deserts and rehabilitation functions in this regard. Several conclusions can be drawn from an examination of Figures 1 and 2.

First, since the utility functions for special deterrence and general deterrence are almost identical in each of these cases, it is impossible to evaluate the influence of these purposes separately. It is conceivable that other situations (e.g., an offender with a bad criminal record) might lead to divergences in these functions, but in the present study it appears that a penalty that is good for one of these purposes is seen as good for the other. The incapacitation function is also always monotonically increasing and thus will also be highly correlated with the deterrence functions, leading to difficulty in separating its effects on appropriateness from those of special and general deterrence. Further, the monotonicity of these functions suggests that none of these three purposes or any weighted average of them could be the sole determinant of the appropriateness of various penalties, since the most severe possible punishment would always be the most appropriate if this were true. In addition, these findings are consistent with earlier findings among both real judges and college students that a concern with deterrence is associated with more severe sentences for all crimes (Hogarth, 1971; McFatter, 1978).

The relation between the rehabilitation and just deserts curves appears to change as crimes increase in seriousness. For the less serious crimes (assault and car theft) the penalties perceived as maximally useful for rehabilitation are more severe than those seen as most deserved, whereas for the serious crimes (rape and murder) the situation is just reversed. It appears that the utility of a penalty for rehabilitation is less sensitive to changes in the seriousness of the crime than is the usefulness of the penalty for providing just deserts. This may provide an ex-

planation for the finding among both judges and college students that for serious crimes a rehabilitation orientation produces less severe sentences, whereas for minor crimes a rehabilitation orientation produces more severe sentences (Hogarth, 1971; McFatter, 1978; Wheeler, Bonacich, Cramer, & Zola, 1968).

Individual analyses. The relative importance weights from Equation 1 were estimated separately for each subject using multiple regression techniques to predict appropriateness ratings from utilities for the individual purposes. Since the utility values for special and general deterrence tended to be very highly collinear, these values were averaged to provide a single function, simply labeled *deterrence*, for each subject. The weight estimates from Equation 1 were obtained by dividing each standardized regression coefficient by the sum of the absolute values of all standardized regression coefficients in the equation, thus normalizing the weights so that their absolute values sum to one for each subject.

A Crime \times Purpose analysis of variance (ANOVA), treating purpose as a repeated measure factor, was carried out on the obtained weights. Table 1 presents the mean weights for each purpose for the four crimes, along with the average squared multiple correlations for subjects' models in each group. Significant effects were found for both purpose, $F(3, 204) = 20.97, p < .001$, and the Purpose \times Crime interaction, $F(9, 204) = 3.91, p < .001$. Both effects were also significant at the .001 and .05 levels using the conservative Greenhouse-Geisser test (Winer, 1971, p. 523-524). Thus, there appear to be differences in the patterns of weights for the various crimes. In three of the four crimes (assault, car theft, and rape) just deserts received the highest average weight, and in the fourth (murder) received the second highest weight. The usefulness of a penalty for giving the offender what he deserves thus appears to be a major determinant of the perceived appropriateness of that penalty. The utility of penalties for rehabilitation also received substantial weight for three of the four crimes and received the most weight on the average for the murder.

It is interesting that the rape is the only

Table 1
Average Relative Importance Weights Calculated for Individual Subjects

Crime	Incapacitation	Just deserts	Rehabilitation	Deterrence	\bar{R}^2
Assault	-.025	.321	.229	-.107	.74
Car theft	-.081	.360	.222	-.004	.77
Rape	-.055	.344	.062	.239	.83
Murder	.114	.167	.253	.102	.87
<i>M</i>	-.012	.298	.192	.058	.80

Note. The standard error of the cell means in the body of the table (based on the within-subjects residual mean square) is .060.

crime for which deterrence received a substantial weight and is the only crime for which rehabilitative utility did not appear to be an important determinant of perceived appropriateness. Apparently, subjects tended to make their judgments as though it were more important to deter the rapist than to rehabilitate him, whereas just the opposite held for the other crimes.

It should be noted that for most subjects nearly all the variance in appropriateness was accounted for by one or two purposes—usually just deserts and deterrence in the rape case, and just deserts and rehabilitation in the others.

Rated importance of purposes. Subjects' ratings of the importance that should be attached to each of the five purposes in deciding upon a sentence were examined in a Crime \times Purpose ANOVA, treating purpose

as a repeated measure factor. The only significant effect was the purpose main effect, $F(4, 272) = 25.76, p < .001$, which was also significant at the .001 level using the conservative Greenhouse-Geisser test. The crime main effect and the Purpose \times Crime interaction were both nonsignificant, $F(3, 68) = .16$ and $F(12, 272) = 1.28$, indicating that the pattern of importance ratings was similar across crimes. Table 2 shows the average importance ratings for each purpose made by subjects in the present study as well as similar data for comparison derived from Hogarth's (1971) almost identically worded questioning of 68 practicing Canadian judges.

On the average, subjects rated individual deterrence to be the most important consideration in deciding on an appropriate sentence, followed by rehabilitation and just deserts. The average importance rating of

Table 2
Average Importance Ratings of Five Purposes by College Students and Judges and Correlations Between Importance Ratings and Estimated Relative Weights

Purpose	Ratings		Average r between student ratings and relative weights
	Judges ^a	College students ^b	
Incapacitation	1.91	2.15 _a	-.22
Just deserts	1.31	3.07 _c *	.16
Rehabilitation	3.16	3.10 _c	.22
General deterrence	2.40	2.60 _b	-.09
Special deterrence	2.20	3.44 _d *	-.09

Note. College student rating means with identical subscripts were not significantly different at the .05 level, using the Fisher protected t test. Each correlation is an average of four correlations (first converted to Fisher Z 's, then averaged and transformed back to r 's), in turn based on 18 observations.

^a From Hogarth, 1971. $n = 68$.

^b $n = 72$.

* College student ratings were significantly ($p < .01$, by t tests) higher than the Canadian judges' ratings for these purposes.

general deterrence was significantly lower than these, whereas the average rating of incapacitation was significantly lower than all the rest. It is worth noting that the college student subjects rated both just deserts and individual deterrence as significantly more important than did Hogarth's (1971) Canadian judges.

Correspondence between computed and subjective weights. A direct comparison of the results shown in Tables 1 and 2 reveals that although in three of the four crimes just deserts received a substantially higher computed or "objective" relative weight on the average than did any other purpose, this pattern is not reflected in subjects' subjective ratings of the importance that each purpose should receive. Subjects' ratings and the regression results do agree, however, in indicating relatively low importance being assigned to incapacitation.

It is possible to examine in other ways the degree of correspondence between subjects' ratings of the importance of the five purposes and the weights derived from subjects' utility judgments. Correlation coefficients were calculated within each crime group between importance ratings for a given purpose and the weights derived in estimating Equation 1 for that purpose. Table 2 also shows the average within-groups correlation for each purpose. Both general deterrence and special deterrence correlations were obtained by correlating ratings for these purposes with the deterrence weights estimated above. None of these correlations is significantly different from zero. This result implies, for example, that subjects who rated just deserts as highly important were not necessarily the ones with large weights for just deserts from Equation 1.

Another perspective on this issue may be gained by looking at the data another way. Correlations were calculated for each subject between his or her importance ratings for each purpose and the relative weights estimated for Equation 1. (Importance ratings for special and general deterrence were averaged to obtain an importance rating for deterrence. Thus, each correlation was based on four pairs of observations.) These correlations were, of course, highly variable, but when transformed to Fisher Z scores and

averaged, the average Fisher Z , corresponding to a correlation of .32, was significantly different from zero ($p < .005$, two-tailed test), providing evidence for at least some correspondence between subjective and computed weights.

Taken together, these results suggest that subjects tended to underestimate the role that the norm of just deserts plays in determining the perceived appropriateness of sentences. On the other hand, they probably overestimated the role of special deterrence in determining appropriateness, especially for the crimes other than murder. These results will be discussed further in connection with Experiment 2.

Experiment 2

The experiment described above investigated the perceptions and sanctioning norms of college students. Although these results are important in understanding how ordinary citizens think about the punishment of offenders, actual sentencing decisions are almost always made by judges of the federal and state court systems. It is, therefore, important to examine perceptions about the utilities of penalties among judges who routinely make sentencing decisions. Experiment 2 focused on the same issues as Experiment 1 but elicited judgments from state district judges in Texas and Louisiana.

Method

Five district judges in Louisiana and three district judges in Texas agreed to participate in the study. Each judge was given four booklets (one for each crime) identical to the ones used in Experiment 1. The judges were asked to complete the booklets at their convenience and return them to the experimenter. All judges returned the booklets within 2-3 weeks of receiving them. Two of the judges (one from Louisiana and one from Texas) returned responses that were unusable; one judge left pages blank in the booklets and the other wrote comments in the margins of his booklets that indicated that he had rated combinations of penalties rather than individual penalties as the instructions requested. This experiment differed from the earlier one in that each judge made judgments for all four crimes rather than for only one.

Results

Aggregate utility functions. Figures 3 and 4 show the aggregate utility functions

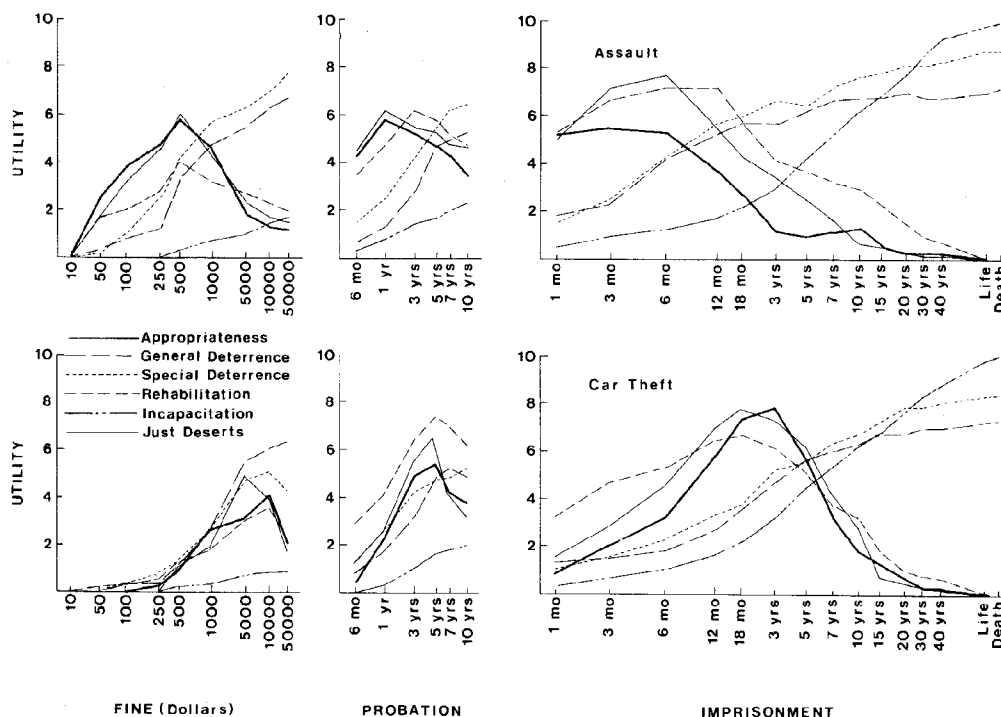


Figure 3. Aggregate utility functions for assault and car theft from district judges' ratings.

derived from the six judges' responses. They are similar in a number of respects to the students' results.

As with the students, the incapacitation, special deterrence, and general deterrence functions are generally monotonically increasing. The just deserts and rehabilitation functions, again, rise to an optimum and then decline, and the overall appropriateness function also exhibits this property. It appears that in the judges' eyes the deterrent utilities of penalties show decreasing marginal returns, especially for the minor crimes. That is, once penalties reach the level of 10 to 15 years imprisonment, substantial increases in sentence severity are not accompanied by strong increases in deterrent utility. Probation terms were generally seen as more useful for rehabilitation than for other purposes, but these judges also saw imprisonment terms as capable of having substantial rehabilitative utility. When fines were seen as useful (which was only for the minor crimes), they were seen as most useful for

just deserts and deterrence rather than for incapacitation or rehabilitation.

Individual analyses. Relative importance weights from Equation 1 were estimated for each judge's responses to each crime in the same manner as in Experiment 1. Again, since special and general deterrence utilities were highly collinear, they were averaged to provide simply a deterrence function. A Judge \times Crime \times Purpose repeated measures ANOVA was carried out on the obtained relative weights. The only significant effect was the purpose main effect, $F(3, 15) = 8.74, p < .005$, which was also significant at the .05 level using the conservative Greenhouse-Geisser test. The mean relative weights for each crime are presented in Table 3 along with the average squared multiple correlations for the judges' models.

It is clear from Table 3 that for the judges just deserts received the largest weight on the average of any of the purposes and this held true for all four crimes. Utility for rehabilitation received the second largest

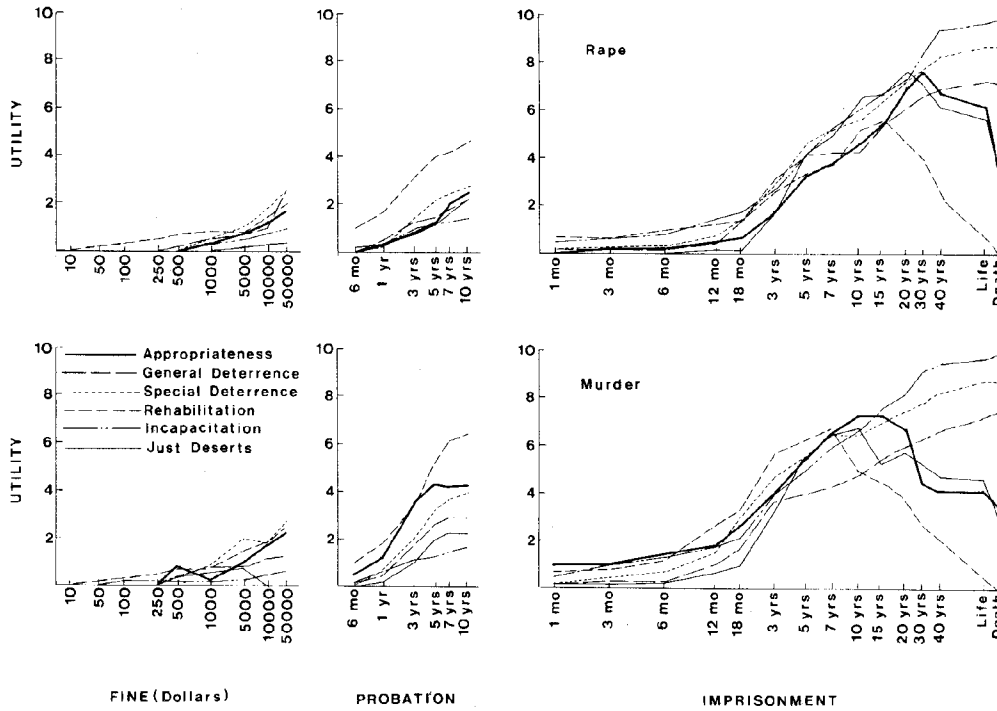


Figure 4. Aggregate utility functions for rape and murder from district judges' ratings.

weight on the average but does not appear to have been anywhere near as influential as just deserts. As the table indicates, the average squared multiple correlations for judges' models that included all four purposes was .83. It is worth noting that if the models are constructed using only just deserts as a predictor of appropriateness the average squared multiple correlation only drops to .73, indicating that utility for just deserts captures most of the variation in appropriateness.

It is of interest to examine the pattern of relative weights estimated for each individual judge. Judging from the degree of controversy surrounding the issue of what purposes ought to be most important, and, indeed, from these judges' responses to that issue reported in the next section, one might expect there to be large individual differences in the pattern of weights estimated for each judge. Table 4 reports the relative weights (averaged over crimes) estimated for each judge.

Although there appears to be some variation in the patterns of weights, for five of the six judges just deserts received the largest weight. Under certain assumptions¹ it is possible to obtain a statistical test of the similarity of the patterns of weights across judges. This is a test of the Judge \times Purpose interaction in the ANOVA. This test yields a nonsignificant result, $F(15, 45) = 1.14, p > .25$, indicating no evidence for reliable dif-

¹There are two ways of justifying this test of the Judge \times Purpose interaction term (see e.g., Myers, 1979): (a) If one is willing to assume that there is no true Judge \times Purpose \times Crime interaction, then the three-way interaction is an estimate of error variance and thus the appropriate error term for testing the Judge \times Purpose interaction. (b) If one treats the four crimes as a random sample from the population of possible crimes of interest, then regardless of whether there is a true Judge \times Purpose \times Crime interaction the three-way interaction is the appropriate error term for the Judge \times Purpose interaction. Neither of these approaches seems totally justifiable; consequently, this significance test should be interpreted cautiously.

Table 3
Average Relative Importance Weights Calculated for Individual Judges in Experiment 2

Crime	Incapacitation	Just deserts	Rehabilitation	Deterrence	\bar{R}^2
Assault	-.028	.400	.182	-.070	.76
Car theft	-.009	.483	.198	.052	.80
Rape	.077	.403	.103	.091	.95
Murder	-.064	.426	.103	.161	.80
<i>M</i>	-.006	.428	.147	.059	.83

ferences in the pattern of weights across judges.

Rated importance of purposes. A Judge \times Crime \times Purpose repeated measures ANOVA was carried out on the judges' ratings of the importance that should be attached to each purpose. There was a significant purpose main effect, $F(4, 20) = 3.59$, $p < .05$, but a test (see Footnote 1) of the Judge \times Purpose interaction was also significant, $F(20, 60) = 3.07$, $p < .01$. This suggests that there are reliable differences among the judges in which purposes they rated as important. Table 4 also shows the average importance rating attached to each purpose by each judge. Judges in Experiment 2 rated just deserts on the average to be the most important consideration followed closely by special deterrence and rehabilitation. As with the college students and Hogarth's (1971) judges, incapacitation was rated of low importance.

Correlation coefficients were calculated for each of the four crimes for each judge

between the relative weights from the judge's model and the importance ratings supplied by the judge for the purposes. (Again, importance ratings for deterrence were obtained by averaging the ratings for special and general deterrence.) Thus each correlation was based on four pairs of numbers, but each judge yielded four correlations, one for each crime. These correlations were converted to Fisher Z scores, averaged for each judge, and converted back to correlations. These average correlations for the six judges were .52, .56, .82, .61, .71, .57. All these correlations are positive and fairly large, suggesting a fair degree of correspondence between how these judges thought the purposes ought to be weighted and how they actually appear to have weighted them. These average correlations are all larger than the corresponding average correlation (.32) for the college students in Experiment 1, suggesting a stronger correspondence between subjective and computed weights for the judges than for the students.

Table 4
Average Estimated Relative Weights and Importance Ratings of the Purposes for Six Judges in Experiment 2

Purpose	Judge						<i>M</i>
	1	2	3	4	5	6	
Average estimated relative weights							
Incapacitation	.17	-.03	-.01	.06	-.19	-.04	-.01
Just deserts	.30	.72	.58	.41	.21	.35	.43
Rehabilitation	.11	.04	.05	.14	.22	.32	.15
Deterrence	-.16	.10	.02	.10	.27	.02	.06
Average importance ratings							
Incapacitation	3.8	2.0	1.5	3.3	1.5	1.5	2.25
Just deserts	3.5	3.0	4.0	4.0	2.8	2.8	3.33
Rehabilitation	3.0	2.8	2.5	2.5	3.3	3.0	2.83
General deterrence	2.8	2.5	2.0	2.8	2.0	3.3	2.54
Special deterrence	3.8	3.5	2.8	3.0	2.5	3.3	3.13

General Discussion

The results of these two experiments present a fairly consistent picture of the perceived usefulness of penalties for the purposes investigated here. Although there appear to be slight differences in details between the utility functions derived from the judges and the college students, the gross characteristics and shapes of the functions are similar for the two groups. One thing that seems clear from the results is that the perceived utility of penalties in serving the purposes can change dramatically from crime to crime. This is especially true (though not surprising) for just deserts. On the other hand, the judged usefulness for incapacitation remains essentially the same across crimes. The perceived deterrent utility of penalties is more interesting when compared across crimes. Figures 3 and 4 suggest that the perceived deterrent utility functions for severe imprisonment terms remain rather constant from crime to crime, whereas the deterrence functions for fines and short imprisonment terms differ markedly from crime to crime. That is, a penalty of 15 years imprisonment is seen as having about as much deterrent utility for rape and murder as it does for assault and car theft. However, a \$1000 fine is seen as much more of a deterrent for the less serious crimes than for the more serious crimes. Why this is so is not clear. Perhaps it reflects a belief that more serious crimes are more difficult to deter.

The rehabilitation function changed from crime to crime for both judges and college students, and neither group hesitated to attribute substantial rehabilitative utility to imprisonment terms. For the judges there was almost always, however, a probation term with nearly as much perceived rehabilitative utility as the maximally rehabilitatively useful imprisonment term. Consistent with this, Hogarth (1971) found that rehabilitation oriented judges in his sample tended to use probation more often than other judges.

If the results of these studies are to be taken at face value, it seems clear that the perceived appropriateness of penalties depends most strongly on their usefulness in giving the offender what he deserves rather than on incapacitative, rehabilitative, or deterrent utility, especially for the practicing

judges. This suggests that an adequate theoretical account of sanctioning norms must devote considerable attention to how people decide what an offender deserves. It also raises a question about the kinds of social research results that are likely to be used by decision makers in this area. Studies of the incapacitative or deterrent effects of actual criminal sanctions on actual crime rates have obvious theoretical and practical interest. It may turn out, however, that even if one were able to pin down precisely the true utility functions for these purposes, perceptions of what appropriate penalties are would not be appreciably affected because incapacitation and deterrence are not the purposes that typically determine perceived appropriateness.

The issue of how ideas about the usefulness of penalties affect judgments about the appropriateness of those penalties is a complex one. There are several things that need to be considered in thinking about this problem. It has been argued that it is important to distinguish between what decision makers say the most important considerations ought to be and how their decisions actually appear to depend on these considerations. It is by no means certain that decision makers will be aware of what factors most strongly influence their judgments. The results shown in Table 4 suggest that, even for the judges (who appeared to have more accurate insight into their policies than the students), there was more disagreement about which purposes ought to be important than there was in the way their judgments appear to have been influenced by those purposes. For five of the six judges, the largest relative weight is associated with just deserts and, in most cases, is very much larger than its nearest competitor. By contrast, only two of the six judges rated just deserts as being the most important.

It is also interesting that there are such apparently large differences between the purposes the judges in Experiment 2 thought should be important and those that Hogarth's (1971) Canadian judges felt should be important. It is, of course, impossible to identify the source of these differences. One could point to changes in the popularity of just deserts and rehabilitation as philosophical justifications of punishment in the intervening 10 years (e.g., von Hirsch, 1976)

or differences in the sociocultural milieu in which the groups of judges operate, and so on, but these are only speculations. The available evidence suggests that a decision maker's expressed punishment philosophy can have significant effects on sentencing judgments (Hogarth, 1971; McFatter, 1978; Wheeler et al., 1968), but it is important to ask whether variations in expressed philosophies have as much to do with perceptions of what penalty is appropriate as does an overriding and common concern with the norm of just deserts. The present results suggest they may not.

In interpreting these results, however, some limitations of the present experiments are worth noting. First, the sample of crimes examined was small, and certain characteristics, such as criminal record, that might influence the use of purposes were eliminated or held constant. Second, the types of crimes themselves may be ones that are likely to call forth retributive reactions (Hall, 1937). Three of the four crimes examined above were crimes against the person and the fourth (car theft) was of such a nature as to constitute the threat of personal harm to innocent bystanders. The strong concern with just deserts evident in the present experiments might be less pronounced if other types of crimes were examined. Even if this were to occur (though it is not at all clear that it would), the apparent underestimation of the influence of just deserts in the present results would still be striking.

Reference Note

1. Wascoe, N. E. *Methods of elicitation of functional relationship curves* (Rep. No. 225). Boulder, Co.; University of Colorado Institute of Behavioral Science, Center for Research on Judgment and Policy, March 1980.

References

- Einhorn, H. J. The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 1970, 73, 221-230.
- Ezorsky, G. *Philosophical perspectives on punishment*. Albany: State University of New York Press, 1972.
- Hall, L. Reduction of criminal sentences on appeal: I. *Columbia Law Review*, 1937, 37, 521-556.
- Hamilton, V. L., & Rytina, S. Social consensus on the norms of justice: Should the punishment fit the crime? *American Journal of Sociology*, 1980, 85, 1117-1144.
- Hammond, K. R., McClelland, G. H., & Mumpower, J. *Human judgment and decision making: Theories, methods, and procedures*. New York: Praeger, 1980.
- Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes*. New York: Academic Press, 1975.
- Hart, H. L. A. *Punishment and responsibility: Essays in the philosophy of law*. London: Oxford University Press, 1968.
- Hogarth, J. *Sentencing as a human process*. Toronto, Canada: University of Toronto Press, 1971.
- Huber, G. P. Multi-attribute utility models: A review of field and field-like studies. *Management Science*, 1974, 20, 1393-1402.
- Keeney, R. L., & Raiffa, H. *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley, 1976.
- McFatter, R. M. Sentencing strategies and justice: Effects of punishment philosophy on sentencing decisions. *Journal of Personality and Social Psychology*, 1978, 36, 1490-1500.
- Morris, N. *The future of imprisonment*. Chicago: University of Chicago Press, 1974.
- Myers, J. *Fundamentals of experimental design* (3rd Ed.). Boston: Allyn & Bacon, 1979.
- Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, 84, 231-259.
- Rossi, P. H., Waite, E., Bose, C. E., & Berk, R. E. The seriousness of crimes: Normative structure and individual differences. *American Sociological Review*, 1974, 39, 224-237.
- Rubin, Z., & Peplau, A. Belief in a just world and reactions to another's lot: A study of participants in the national draft lottery. *Journal of Social Issues*, 1973, 29, 73-93.
- Sebba, L. Some explorations in the scaling of penalties. *Journal of Research in Crime and Delinquency*, 1978, 15, 247-265.
- Sellin, T., & Wolfgang, M. *The measurement of delinquency*. New York: Wiley, 1964.
- Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, 28, 1-39.
- Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, 6, 649-744.
- Vidmar, N. Retributive and utilitarian motives and other correlates of Canadian attitudes toward the death penalty. *Canadian Psychologist*, 1974, 15, 337-356.
- Vidmar, N., & Miller, D. Socialpsychological processes underlying attitudes toward legal punishment. *Law & Society Review*, 1980, 14, 565-602.
- von Hirsch, A. *Doing justice: The choice of punishments* (Report of the Committee for the Study of Incarceration). New York: Hill & Wang, 1976.
- Wheeler, S., Bonacich, E., Cramer, M. R., & Zola, I. K. Agents of delinquency control: A comparative analysis. In S. Wheeler (Ed.), *Controlling delinquents*. New York: Wiley, 1968.
- Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Received August 3, 1981 ■