Florida Law Review

Volume 18 | Issue 4 Article 3

March 1966

Optimizing Legal Policy

Stuart S. Nagel

OPTIMIZING LEGAL POLICY

STUART S. NAGEL*

Various scholars have written in recent years about the quantitative optimization of managerial decisions[1] and a few have dealt with social decisions.[2] It is the purpose of this article to offer an alternative quantitative method to the approaches given in those materials. This alternative method is designed to be particularly applicable to proposed legislation and, to a lesser extent, to judge-made law. It emphasizes estimated correlation coefficients to determine the relation of policies to goals and paired comparisons to weight goals and to determine the likelihood that a policy will be adopted. In order to make the method simpler for the reader to perceive, elementary algebraic symbols and dichotomous variables will be used throughout the article. No knowledge beyond high school algebra will be presumed. Two previous studies provide lengthy specific applications of many of the broader methods and concepts presented here.[3]

---

*B.S. 1957, LL.B. 1958, Ph.D. 1961, Northwestern University; Associate Professor of Political Science, University of Illinois; Member of the Illinois Bar; Presently on leave serving as an Assistant Counsel to the Senate Subcommittee on Administrative Practice and Procedure of the Committee on the Judiciary.

The author gratefully acknowledges the perceptive critique of this paper, which was offered by Mary Ellen Caldwell, Layman Allen, and Myres McDougal of the Yale Law School and Thomas Cowan of the Rutgers Law School.

1. See, *e.g.*, CHURCHMAN, ACKOFF & ARNOFF, INTRODUCTION TO OPERATIONS RESEARCH (1957); MILLER & STARR, EXECUTIVE DECISIONS AND OPERATIONS RESEARCH (1960); SCIENTIFIC DECISION MAKING IN BUSINESS—READINGS IN OPERATIONS RESEARCH FOR NON-MATHEMATICIANS (Shuchman ed. 1963). Thomas Cowan has recently called the attention of legal theorists to operations research materials. Cowan, *Decision Theory in Law, Science, and Technology*, 17 RUTGERS L. REV. 499 (1963).

2. ARROW, SOCIAL CHOICE AND INDIVIDUAL VALUES (2d ed. 1963); BRAITHWAITE, THEORY OF GAMES AS A TOOL FOR THE MORAL PHILOSOPHER (1955); BRAYBROOKE & LINDBLOM, A STRATEGY OF DECISION — POLICY EVALUATION AS A SOCIAL PROCESS (1963). Although the Braybrooke & Lindblom book is not quantitative, the authors do recognize "the boundary between problems that can be solved by calculation and those that must be treated by a strategy of multiple adjustment has thus shifted. With the development of more and more sophisticated computer techniques, the boundary may be expected to go on shifting." *Id.* at 247. The policy-making scheme of this article mainly involves policies that promote a medium degree of change and are based on a medium degree of understanding. *Id.* at 78.

3. Nagel, *Testing the Effects of Excluding Illegally Seized Evidence*, 1965 WIS. L. REV. 283 (particularly Figure 1); Nagel, *Simplified Bipartisan Computer Redistricting*, 17 STAN. L. REV. 863 (1965).

One Policy and One Goal

Correlation Coefficients

The simplest policy problem merely involves attempting to determine whether a given policy or means to an end will increase or decrease the presence of a given intermediate or ultimate goal. An X can be used as a symbol for a policy, and a Y can be used as a symbol for a goal. In terms of correlation coefficients, this problem involves determining whether there is a positive (direct), negative (inverse), or zero correlation between X and Y.

To determine the direction of the correlation between X and Y, one can compare a group of geographical units or individuals who have been subjected to X with a group of geographical units or individuals who have not been subjected to X or who have been subjected to less X. The first group can be referred to as group E for experimental, and the second group as group C for control. The geographical units can be communities, states, countries, or other units, and the individuals can be businessmen, school children, taxpayers, or other individuals depending on whom the policy is supposed to affect. The entities may also be time periods or events some of which had X present and some of which had X absent. If the proportion of group E that is high on Y is a greater proportion than the proportion of group C that is high on Y, then there is a positive correlation between X and Y; and if all other things are temporarily held equal or constant for the sake of discussion, then X should be adopted. If the relative size of these proportions (that is, P_E and P_C) is reversed, then there is a negative correlation between X and Y; and X should be rejected.[4] The difference between these two proportions represents an approximate measure of the degree of correlation between X and Y. It is symbolized r and can range from -1.00 to 0 to $+1.00$. The relationships discussed in this paragraph can be shown in a four-cell table like Table I.[5]

In correlating X and Y, one should try to have entities in group E that are like the entities in group C with regard to any Z characteristics (that is, region, prosperity, or industrialism) that cause X and also cause Y so as to eliminate the possibility of spurious correlation between X and Y. E and C, of course should be unlike on X

---

4. If Y can be measured in degrees rather than merely as high and low, then one would determine whether the entities in the E group had a higher average Y than the entities in the C group.

5. For further detail on gathering empirical data and correlating variables in legal research see Nagel, *Testing Empirical Generalizations in Legal Research*, 15 J. LEGAL ED. 365 (1963) and the references cited therein.

Table I. Correlating Policy X with Goal Y (dichotomous variables)

	Group C $-X$	Group E $+X$	
$+Y$	N_1	N_3	$N_1 + N_3$
$-Y$	N_2	N_4	$N_2 + N_4$
	$N_1 + N_2$	$+ \quad N_3 + N_4$	$= N$ entities

$P_C = N_1/(N_1 + N_2)$ $\qquad$ $r = P_E - P_C$

$P_E = N_3/(N_3 + N_4)$ $\qquad$ $\text{chi}^2 = r^2 N$

$M_Y = (N_1 + N_3)/N$ $\qquad$ $S_Y = \sqrt{(M_Y)\ (1 - M_Y)}$

$M_X = (N_3 + N_4)/N$ $\qquad$ $S_X = \sqrt{(M_X)\ (1 - M_X)}$

and the choosing of entities should be unrelated to how they are positioned on Y. One can also apply (1) a chi-square test to determine whether the r is big enough given the N to be considered not due to chance and (2) a test of reasonableness in light of known relationships. A correlation coefficient can be thought of as being related to the probability that X will achieve Y. It would, however, be erroneous to substitute P_E alone for r because P_E merely indicates the probability of $+Y$ occurring if $+X$ occurs. P_E does not indicate the probability of $+Y$ occurring even if $-X$ occurs. The r takes both P_E and P_C into consideration. As an alternative to determining at one point in time whether each entity in group E and group C was high or low on Y, one can determine whether each entity in group E and group C underwent a big or a little increase in Y after the entities in E were subjected to X.

Relevant data can frequently be obtained from available records or from questionnaires. Most X's and most Y's though cannot be measured like physical or even monetary things can, but entities can at least be positioned as having X or Y present or absent, as high or low, or as more or less than some entity that is used as a standard. It may, however, sometimes be impractical to calculate r because of lack of available data, lack of time or money to obtain or simulate data, or because of the presence of an X variable that cannot be simulated and is practically unique in world history. In such a situation, one will still gain insights into the relation between X and Y by attempting to roughly estimate what P_E and P_C might be or by attempting to have others make such an estimation. The estimate may be based on analogies, known relationships between other variables, knowledge of human nature, or an informed imagination. In making such an estimate one can simplify the arithmetic by taking an N of 20 and estimating what quantity of the 20 are likely to be $+X$ and what quantity are likely to be $+Y$ and then estimating what quantity of the $+X$ entities are also likely to be $+Y$. Once these three estimates are made, all the other values shown in Table I can be calculated by simple arithmetic. One can assume that $N_1 + N_2 = N_3 + N_4$ and that $N_1 + N_3 = N_2 + N_4$ if no better estimate can be made. Where policy decisions have to be made, a rough estimate of the relation between X and Y may be better than no estimate and no decision.

Although an X having a calculated or estimated negative correlation with Y should generally be rejected, it may be wise to adopt X if only a small negative correlation is involved and if there is more to be lost by an error of rejecting an X that really had a positive correlation with Y than by an error of adopting an X that really had a negative correlation.

Regression Equations

Suppose one finds or estimates a positive correlation between policy X_1 and Y and also a positive correlation between policy X_2 and Y and one has the resources to adopt both X_1 and X_2, should he then adopt both policies? Not necessarily. Such a pair of correlations merely indicates X_1 alone will increase Y, and X_2 alone will increase Y. If the combined effect of X_1 and X_2 does not produce sufficient Y, then a policy of X_3 may have to be sought. On the other hand, if policy X_1 alone or policy X_2 alone can produce sufficient Y, then it might be wasteful to adopt both policies. This problem involves finding an equation that indicates how many units of X are needed to produce a given quantity of Y.

Such an equation (called a regression equation) can be symbolized as follows: $Y = a + bX$, where a is referred to as the a-coefficient and b is referred to as the b-weight. The a can be calculated by the formula, $a = M_Y - M_X b$, where M_Y is the mean of Y and M_X is the mean of X. The b can be calculated by the formula, $b = r(S_Y/S_X)$, where S_Y is the standard deviation of Y and S_X is the standard deviation of X. Simple formulas for calculating M's and S's for a four-cell table are given in Table I. Standard statistics textbooks give more complex formulas for calculating M's, S's, and r where X and Y have not both been dichotomized. In using the regression equation where Y has been dichotomized, simply substitute a 1 for $+Y$ and solve for X. If X is also dichotomized, then an X value equal to or greater than .5 equals $+X$ and an X value less than .5 equals $-X$. Likewise one could solve for Y by substituting a 1 or zero for X in the equation. If r is positive, then $+X$ will yield $+Y$; and if r is negative, $-X$ will yield $+Y$. Where X or Y are continuum variables rather than dichotomies, more precise values than 1's or zero's can be substituted for X and Y and the regression analysis becomes more worthwhile.

The regression equation given presupposes a roughly linear relationship between X and Y. In other words the more X one adopts, the more Y one will get although additional units of X may not produce proportionately additional units of Y. In a completely curvilinear relationship, the more X one adopts the more Y one will get up to a maximum point, and then additional units of X take away units of Y rather than merely add smaller units of Y. At least a six-cell table like Table II is needed to reveal such a curvilinear correlation. If P_2 is greater than P_1 and greater than P_3 a curvilinear relation is present. The regression formula for such a hill-shaped relation is $Y = a + b_1 X + b_2X^2$. To calculate a, b_1, and b_2 requires solving a system of three simultaneous equations. In a 3-by-2 table like Table II, however, when an entity is in the first category on X, one can predict P_1 on Y; when an entity is in the second category on X, one can predict P_2 on Y; and when an entity is in the third category, one can predict P_3 on Y. One can make similar predictions on Y for each category of X regardless of the number of X categories so long as Y is dichotomized. This gives the same prediction as the regression equation does.[6]

Table II. A Curvilinear Relation Between X and Y.
(P_2 greater than P_1 or P_3)

	low X	middle X	high X
$+Y$	N_1	N_3	N_5
$-Y$	N_2	N_4	N_6

$P_1 = N_1/(N_1 + N_2)$. $P_2 = N_3/(N_3 + N_4)$. $P_3 = N_5/(N_5 + N_6)$.

6. For further detail on linear and curvilinear regression with one or multiple X's see BLALOCK, SOCIAL STATISTICS 273-358 (1960); PETERS & VAN VOORHIS, STATISTICAL PROCEDURES AND THEIR MATHEMATICAL BASES 425-35 (1960).

Interpretation of the data in a table like Table I or II will inform the policymaker whether adopting or adding to X will increase or decrease Y. That would be an inductive approach. A deductive approach, on the other hand, involves deducing from a regression or other equation whether or not X should be adopted. The equation itself may be derived by induction from the analysis of data or by deduction from related empirically-tested formulas, intuitively accepted axioms, or from definitions. If the regression equation involves a hill-shaped relation and both X and Y can be numerically measured, then one can use simple calculus to determine the number of units of X needed to reach a peak on Y before diminishing total returns set in. To do so, one determines the derivative of Y with respect to X given the regression equation, replaces the derivative sign with a zero, and then solves the resulting equation for X. This would be one of the rare instances where calculus, as contrasted to algebra or statistical analysis, has any direct value to legal research.[7]

MORE THAN ONE POLICY OR GOAL

Multiple Policies and One Goal

Suppose one wishes to compare two or more policies (X_1, X_2, X_3, et cetera) that have been proposed for achieving desired goal Y. In terms of correlation coefficients, the logical thing to do is to compare the correlation coefficient (r) between X_1 and Y with the coefficient between X_2 and Y, and so on down to the coefficient between X_n and Y where X_n is the last X. If other considerations are held constant, then the X with the highest r is best, the X with the next highest r is next best, and so on. X's with positive r's are worthy of adoption assuming the X's are not mutually exclusive, and X's with negative r's are not. As previously mentioned, if a certain correlation coefficient cannot be calculated because of lack of data, it may be better to estimate it or have others who are knowledgeable estimate it than to abandon the correlation approach. Regression equations can be calculated for multiple X's by a method analogous to the method described where one X was involved. The general regression formula is $Y = a + (b_1X_1) + (b_2X_2) + \dots + (b_nX_n)$. Standard computer programs make such calculations relatively easy after the individuals, geographical units, or other entities involved have been categorized on the X and Y variables.

7. The derivative of a hill-shaped equation $Y = a + b_1 X + b_2X^2$ equals $b_1 + 2b_2X$. PROTTER & MORREY, COLLEGE CALCULUS WITH ANALYTIC GEOMETRY 126-27, 160 (1964). Thus when the derivative sign is replaced with a zero, the X that will give a maximum Y equals $-b_1/2b_2$.

Where there are a great many mutually exclusive *X*'s, a computer can also quickly determine the *Y* that all or many of the *X*'s will produce. Such a situation is involved in determining the optimum policy for assigning counties or other geographical units to legislative districts. The *Y* or *Y*'s in redistricting are mainly equality and compactness.[8]

One Policy and Multiple Goals

The problem is a little more complicated if one has many *Y*'s in mind and one wants to know whether or not to adopt policy *X*. If all the *Y*'s are equally worthy and nonduplicative, then in terms of correlation coefficients, one would logically add the correlation coefficient between X and Y_1 (symbolized r_{Y_1X}) to r_{Y_2X} to r_{Y_3X} and so on. If the sum of the *r*'s is plus or positive and the other considerations are held constant, then *X* should be adopted because a positive sum indicates that *X* correlates positively with the desired goals more than it correlates negatively. If the sum of the *r*'s is minus or negative, then *X* should be rejected for the converse reason.

If, however, all the *Y*'s are not equally valuable as is more likely to be the case, then each r_{YX} should be multiplied by the weight (symbolized w) or relative worth of the X in the r_{YX} before summing the r_{YX}'s. How can one assign weights to each goal? A simple and meaningful method is the method of paired comparisons. To apply this method, one pairs each goal with each other goal. If there are N goals, then there are $N(N - 1)/2$ comparisons. In preparing the list of pairs, every goal should appear sometimes on the left and sometimes on the right, and no goal should be involved in two successive pairs. The list of pairs is then given to a group of legislators, philosophers, public opinion experts, or some segment of the public depending on whose values one is interested in. Each person in this group should indicate which goal he prefers in each pair or which goal he thinks the public would prefer. A matrix is then prepared in which one lists each goal along both the top and the side. The cells in the matrix indicate the proportion of times the goal at the top was chosen in preference to the goal at the side. The sum of these proportions downward for each goal can represent the value points or weight for each goal, or one can mathematically manipulate the matrix to determine more precise value points and intervals between goals. Paired comparison programs, as well as correlation and regression programs, are now available in various computer program libraries.[9]

8. Nagel, *Simplified Bipartisan Computer Redistricting*, note 3 *supra*.

9. STATISTICAL SERVICE UNIT OF THE UNIVERSITY OF ILLINOIS, MANUAL OF COM-

The number of comparisons becomes unwieldly if the number of goals exceeds 10, but that is unlikely to occur in most policy problems. If it does occur, then as an alternative to the paired comparisons method, one can use the method of successive categories to weight the goals. This method involves asking the evaluators to place a goal in one of five or more categories such as (1) highly undesirable, (2) mildly undesirable, (3) neither desirable nor undesirable, (4) mildly desirable, (5) highly desirable. A matrix is then prepared with the goals on the side and the successive categories on the top. The cells in the matrix indicate the proportion of times the goal at the side was placed in the category at the top. The matrix is then mathematically manipulated to assign each goal a definite weight.

Neither the paired comparisons method of weighting nor the successive categories method adequately indicates the degree of difference between the Y's if all the evaluators always prefer a certain Y or always reject a certain Y, or if the evaluators always put a certain Y in the highest category or always in the lowest category. This problem can be handled by increasing the number of categories or by providing a thermometer-type rating scale on which the evaluators can attempt to position the goals. Such ratings, however, can be checked by each evaluator for consistency by seeing whether the combined weights of various preferred combinations total more than the combined weights of various nonpreferred combinations. The only such comparisons worth making are those where the outcome is not obvious on the basis of the ranks alone. For example, where there are four separate, compatible goals rated in descending order Y_1, Y_2, Y_3, and Y_4, which combination is more valued Y_1 or $(Y_2 + Y_3 + Y_4)$; $(Y_1 + Y_4)$ or $(Y_2 + Y_3)$; Y_1 or $(Y_2 + Y_3)$; Y_1 or $(Y_2 + Y_4)$; or Y_1 or $(Y_3 + Y_4)$; Y_2 or $(Y_3 + Y_4)$? If the sums of the weights are inconsistent with any of these value judgments, then the evaluator can adjust the weights to conform with the judgments. The weights assigned by the evaluators to each Y can then be averaged to give average weights for the group of evaluators.[10]

Multiple Policies and Multiple Goals

Suppose one wishes to compare two or more policies and he has many goals in mind. This is the problem of multiple X's and mul-

PUTER PROGRAMS FOR STATISTICAL ANALYSIS (1964) [hereinafter cited as SSU MANUAL].

10. For further detail on the method of paired comparisons, successive categories, and other methods for preparing an evaluation or prediction scale see GUILFORD, PSYCHOMETRIC METHODS 154-301 (2d ed. 1954). For alternative methods not covered in GUILFORD see CHURCHMAN, ACKOFF & ARNOFF, *op. cit. supra* note 1, at 136-54.

tiple Y's simultaneously. The general formula for the relative utility (symbolized U) of X_1 or any X given any number of Y's is: $U_{X_1} = (r_1w_1) + (r_2w_2) + \ldots + (r_nw_n)$, where r_n is a shortened way of writing $r_{Y_nX_1}$ (that is, the correlation between X_1 and Y_n), and w_n is a shortened way of writing w_{Y_n} (that is, the weight of Y_n). Thus if other considerations are held constant, then the X with the highest U is best, the X with the next highest U is next best, and so on. X's with positive U's are worthy of adoption, and X's with negative U's are not. For example, if one is attempting to decide between X_1 (which produces $+Y_1$ and $-Y_2$) as contrasted to X_2 (which produces $-Y_1$ and $+Y_2$), then one should logically choose the X that produces the Y with the greater w unless the correlations offset the difference in w's. In other words, if only one X can be chosen, choose the X with the greater U. If, however, the U_{X_1} is only Q times the value of U_{X_2} but one can obtain more than Q times as much X_2 than he can obtain X_1, then one should choose the maximum X_2 obtainable all other things held constant.

X_1 may have a higher U than X_2 if the future is a period of $+Z$, but a lower U than X_2 if the future is a period of $-Z$. If it is practically certain that the future will be $+Z$ then X_1 is clearly better than X_2. If, however, the estimated probability that $+Z$ will occur is P and the ratio $(1 - P)/P$ is more than the ratio U_{X_1}/U_{X_2}, then X_2 should be preferred. In other words, where the U of an X is contingent on the occurrence of some future condition, the U should be multiplied by the probability of the condition occurring when comparing the U with the U's of other X's. If it is impossible to arrive at a roughly meaningful probability that a crucial Z will occur, then it is traditional to assume that Z has a .50 chance of occurring.[11]

The utility score in the above formula is only a relative score. It enables one to rank policy proposals in light of the goals of the policymaker or interest group concerned. To determine the quantity of goal achievement that a given set of policies will achieve when one has a number of goals in mind, one can calculate the *a*-coefficient and *b*-weights for a regression equation for each goal using the general regression formula previously given. Then substitute the quantity or category of $X_1, X_2, \ldots,$ and X_n to be adopted into each regression equation, and solve for Y in each equation. The set of Y values obtained represents the total quantity of goal achievement predicted. The separate Y values, however, cannot be added together since that would involve adding different kinds of things together. On the other hand, one can talk about maximizing the over-all goal (G) where G

11. MILLER & STARR, *op. cit. supra* note 1, at 79-100.

equals the product of the Y values with each Y value raised to the power of its weight.[12] In other words, $G = (Y_1^{w_1}) (Y_2^{w_n})$. A correlation coefficient between a set of X's on the one hand and a set of Y's on the other hand can be determined by a technique known as "canonical correlation."[13]

To maximize Y (rather than obtain a specific Y) where there is a linear relationship between X and Y, choose the maximum X available. To maximize Y where there is a hill-shaped relationship between X and Y, choose the X equal to $-b_1/2b_2$ as previously mentioned in note 7 *supra*. Where there is more than one X, the optimum value for each X depends on the cost of one unit of X and on the maximum amount of money available to achieve Y. Suppose X_1 costs 10,000 dollars per unit, X_2 costs 20,000 dollars per unit, and there is a maximum of approximately 500,000 dollars to spend in order to achieve Y. In such a situation the optimizing problem becomes one of maximizing Y where $Y = a + b_1X_1 + b_2X_2$ and where $10X_1 + 20X_2$ is less than or equal to 500. If one has the data to solve a, b_1, and b_2 in the regression equation, then this regression equation and the above cost equation can be put into a programming routine that will indicate the appropriate units of X_1 and the appropriate units of X_2 in order to maximize Y in light of these two equations.[14] One might also wish to add an equation that X_2 be less than or equal to $-b_1/2b_2$. The regression equation would then be $Y = a + (b_1X_1) + (b_2X_2) + (b_3X_2^2)$ where the third and fourth terms indicate X_2 and Y have a hill-shaped relation.

Listing Policies and Goals

In order to make the analysis more meaningful where one is working with more than one X, he should be careful to eliminate from consideration those X's that are duplicative. A duplicative X is one that represents basically the same thing as another X although the two X's are described in different words. Likewise duplicative Y's should be eliminated where one is working with more than one Y. In instances where one can obtain or estimate a meaningful correlation coefficient for every X with every other X from tables like Table I, one can put these coefficients into a correlation matrix and

12. *Id.* at 161-65. Where the weights are decimals, large numbers, or negative numbers one can calculate such a G easier by the formula: $\text{Log } G = (w_1 \text{ Log } Y_1) + (w_2 \text{ Log } Y_2) + \ldots + (w_n \text{ Log } Y_n)$. RICHARDSON, COLLEGE ALGEBRA 391-404 (1958).

13. TINTNER, ECONOMETRICS 114-21 (1952); SSU MANUAL.

14. See the linear programming routine described in the SSU MANUAL and in KEMENY, SNELL & THOMPSON, INTRODUCTION TO FINITE MATHEMATICS 249-65 (1957).

factor the matrix thereby giving a set of nonoverlapping X factors smaller in quantity than the original X's. These X factors can then be correlated with the Y's. Likewise one could conceivably reduce the number of Y's and eliminate overlapping Y's by intercorrelating each Y with each other Y and then factoring the resulting correlation matrix.[15] It is, however, generally easier and more meaningful to reduce the X's and Y's by doing some hard thinking and rewording with regard to the distinctiveness of each X from each other X and each Y from each other Y. Careful thinking will also help to eliminate impractical X's. It might also be noted that where there are multiple X's, the b-weight for each X is generally computed in such a way as to hold constant statistically the other X's. Thus where such b-weights are available, they can be substituted for the r's and then the proper U formula requires multiplication of powers rather than the addition of products since the b's are not pure numbers like the r's.[16] The U formula under such circumstances becomes

$$U_{X_1} = (b_1^{w_1})\ (b_2^{w_2}) \ldots (b_n^{w_n}).$$

In order to do a factor analysis or calculate a multiple regression equation with b-weights that hold other X's constant, it is necessary to correlate each X with each other X. Such correlations may also be helpful by themselves in eliminating duplicate X's or Y's since duplicates will have a correlation of approximately $+1.00$ with each other although a perfect correlation may be due to causes other than duplication. The correlation between two X's or two Y's that are mutually exclusive is -1.00. The correlation between two X's or two Y's that are not duplicative or mutually exclusive can be calculated by gathering or estimating data for a table like Table I. If this is impractical, it may be reasonable to say there is a zero correlation between the X's or the Y's in which case P_1 equals P_2.

In listing the goals involved, one should include possible side effects, both desired and undesired, as well as intended goals. In the calculations, however, side effects need not include effects of effects of the policies otherwise the evaluation process might become unwieldy. One of the Y goals should generally be low monetary cost, and the relation between the X's and such a Y can be determined by accounting and budgeting techniques rather than by the gathering of behavioral data. For ease in handling, all goals should generally be stated in an affirmative way so that $+Y$ is what is desired and $-Y$ is what is not desired and all w's are thus positive. In wording the goals, one should try to be as explicit as possible. A well-stated set

15. Factor analysis is discussed in FRUCHTER, INTRODUCTION TO FACTOR ANALYSIS (1954) and is programmed in the SSU MANUAL.

16. See the references cited in note 12 *supra*.

of goals should possibly consider the intensity sought, the coverage over persons, and the coverage over time. There is, however, no quantitative system for making a best list of goals as contrasted to weighting the goals within a list or as contrasted to optimizing a set of policies for achieving goals. One can, though, determine whether a goal should be adopted if the goal is a policy toward a higher goal. Ultimately, however, one has to resort to pure value judgments that cannot be correlated with a higher goal when there is no higher goal to use as a criterion. Likewise, although policies can be quantitatively tested and ranked, there is no quantitative method for generating policy ideas. Such ideas are largely dependent on awareness of the relevant literature, on imaginative creativity, and on trial and error.

THE PROBABILITY OF POLICY ADOPTION

Determining the probability (symbolized P) that a given policy will be adopted is important because regardless whether the utility of a given policy is positive, it would generally be wasteful of resources for an interest group outside or inside the membership of a legislature or a judicial body to push the adoption of a policy that has a low probability of being adopted, unless there is something to be gained (such as publicity) even if the policy does not pass, or unless the policy is being pushed with the intention of compromising on a milder passable policy.[17] Likewise it is wasteful of the time, money, and friendship resources of an interest group to push policy X, thinking this is the best they could hope for when with just a little extra effort they could have gained the adoption of both X_1 and X_2, unless there is something to be gained by passing a weak policy by a large majority (such as significantly increased enforcement and compliance) rather than a stronger policy by a narrower majority. In this context P equals the proportion of the policymakers that are likely to vote in favor of the X under consideration. If P is considerably less than 51 per cent, it may as well be zero.

The logical ideal strategy for a policymaker or interest group seems to be to push the policy or combination of policies that has the greatest relative utility (U) or goal achievement (Y or G) to the policymaker or the interest group involved, provided the policy has at least a .51 probability of being adopted. In practice, a policy-

17. The U of an X does not equal Y times w times P because such a formula falsely assumes that if Y_{X_1} / Y_{X_2} is greater than P_{X_2} / P_{X_1}, then X_1, which has a highly unlikely chance of being adopted should be pushed in Congress or the Supreme Court in preference to X_2, which has a good chance of being adopted. Clearly a small Y achievement is better than none at all.

maker may rationally push a desired policy for nonpublicity, nonbargaining motives even though it appears to have only a .40 probability of adoption in order to avoid an error of underestimation which would result in a bill not being pushed which would have been adopted. Likewise a policymaker may rationally push X_1 and not X_1 plus X_2 combined even though X_1 plus X_2 appears to have a .60 probability (and both X's are desired by the policymaker) in order to avoid an error of overestimation which could result in the loss of both X_1 and X_2.

How does one calculate the probability that a policy will be adopted? A policy or policies can be given a probability score by first putting it or them on a list along with policies that have been recently adopted or rejected by the current legislature or court involved. The list can then be submitted in the form of paired comparisons (pairing each policy with each other policy) to a group of persons who have some knowledge of the attitudes of the policymaking body. Each person should indicate which policy in each pair would be more likely to be passed. Using the method of analysis previously mentioned for paired comparisons, each policy can be given an adoption score. To give the proposed policy a specific probability, calculate the average of the per cent of favorable votes favoring the policy immediately below and immediately above the proposed policy on the adoption scale. If the number of paired comparisons becomes unwieldy the method of successive categories of adoption-probability previously mentioned can be used. If one believes that he is an adequate legislative predictor or goal evaluator, then he can assign rough adoption probabilities to the proposals and weights or ranks to the goals without using the pairing or categories method and without using other persons' judgments.

For example, if X has a paired-comparison adoption score of 4.9 and a bill with an adoption score of 4.4 was approved by the United States Senate by a 55-to-45 vote (55 per cent favorable), and a bill with an adoption score of 5.7 was approved by the United States Senate by a 70 to 20 vote (79 per cent favorable), then one can say that the probability of X passing is .67. Instead of simply averaging .55 and .79 one could solve for P by interpolation where $(P - .55)/(.79 - .55) = (4.9 - 4.4)/(5.7 - 4.4)$. A probability could be similarly calculated for the House of Representatives, and then the two probabilities could be averaged to give an over-all probability of congressional adoption.

It would be less meaningful to attempt to arrive at a probability of adoption by analyzing the correlation coefficients between the presence or absence of certain factual elements surrounding the policy and the occurrence or nonoccurrence of adoption using past policy proposals as entities. Such an approach, although possibly useful in predicting adjudication outcomes, does not adequately consider (1) the significant effects of changes in policymaking personnel as contrasted to policy-applying personnel, (2) the relative uniqueness of policymaking proposals as contrasted to case adjudications, and (3) the insights of knowledgeable persons. Likewise any method that attempts to position the attitude of each individual policymaker in a 535-man legislature would probably be too unfeasible to be useful, although sometimes an inside political leader can make an accurate pre-vote survey of nearly all the members of a legislature.

There is no necessary relation between the utility of a policy and its probability of adoption if utility is determined in terms of the values of specific policymakers or interest groups. If, however, a policy has a high utility in terms of the values of the general public and the policymakers are aware of this utility, then, in a democratic society, in the long run the policy will probably be adopted. In the short run though, there may be a big difference between general utility and adoption probability given such institutions as the gerrymander, filibuster, rules committee veto, and Negro voter deprivation as well as lack of accurate information on relations between policies and goals.

CONCLUSIONS

The key principles in this optimizing scheme in abbreviated form are: (1) the relative utility of a policy proposal equals the sum of the correlation-weights for each goal relevant to the policy; (2) the goal achievement of a set of policy proposals equals the sum of the *a*-coefficient and the *b*-weights times the *X* values for each policy; and (3) the policy or combination of policies that should be pushed is the one that has the greatest relative utility or goal achievement with the general proviso that the policy have approximately a .51 probability of being adopted. The most important formulas are

thus: $U_{X_1} = (r_1w_1) + (r_2w_2) + \dots + (r_nw_n)$; and

$$Y = a + (b_1X_1) + (b_2X_2) + \dots + (b_nX_n).$$

The method presented for optimizing legal policy is definitely not meant to be one that can always be applied. It is meant to be a method of thinking whose application is possibly worth striving for wherever it can be applied in part or in whole, even if some of the components have to be estimated rather than calculated precisely. The method presented is also hopefully meant to stimulate further analysis of the applications of operations research and statistical analysis to optimizing legal policy.