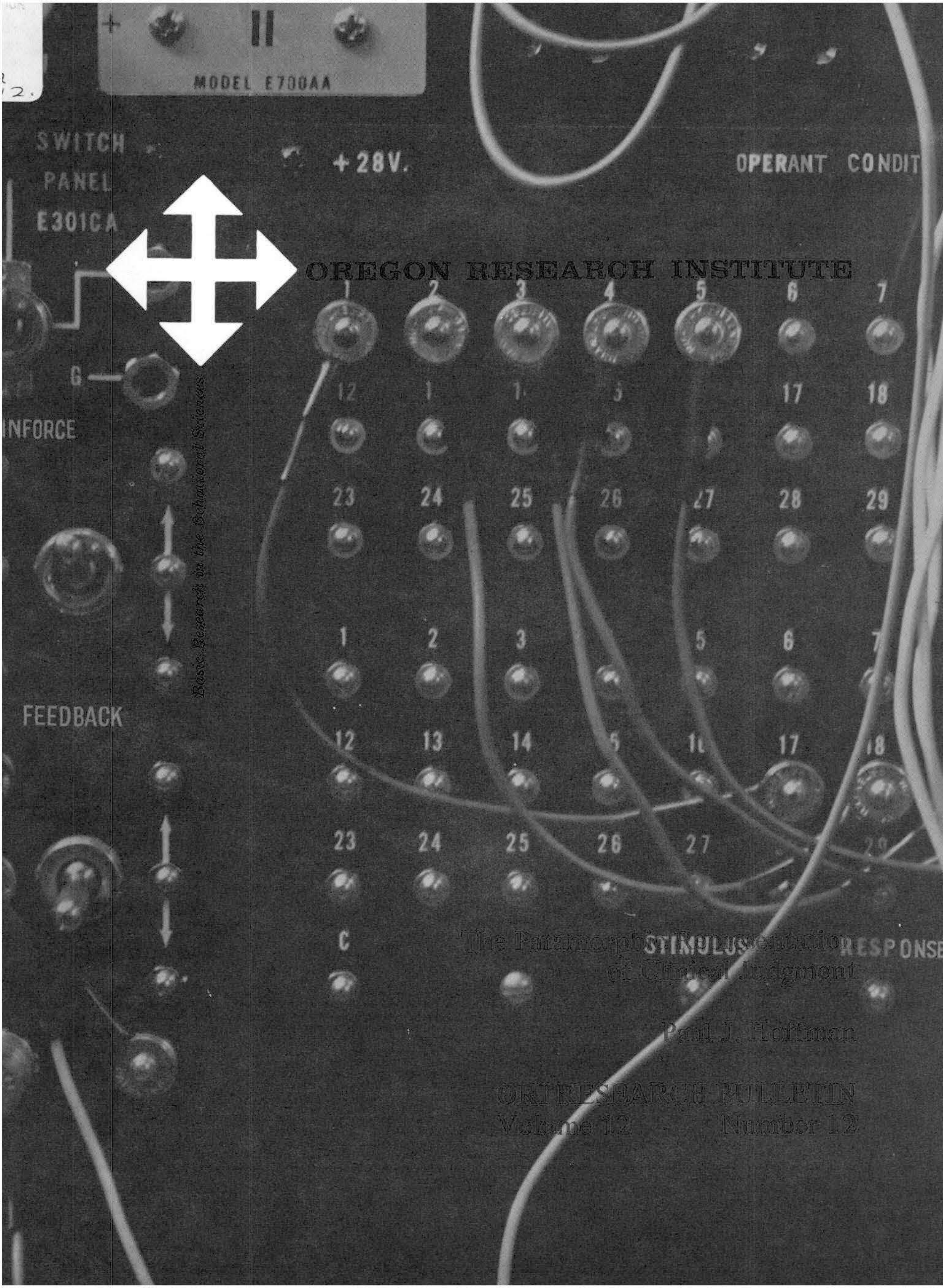

OREGON RESEARCH INSTITUTE

Basic Research in the Behavioral Sciences

The Paramorphic Representation of Clinical Judgment

Paul J. Hoffman

ORI RESEARCH BULLETIN
Volume 12 Number 12

OREGON RESEARCH INSTITUTE

The Paramorphic Representation of Clinical Judgment

Paul J. Hoffman

Oregon Research Institute

Research Bulletin
Vol. 12 No. 12

THE PARAMORPHIC REPRESENTATION OF CLINICAL JUDGMENT[1]

Paul J. Hoffman

The primary task of clinical diagnosis is that of collecting, evaluating, and assimilating information with respect to the patient. The starting point is the information itself; this may be in the form of laboratory test results, biographical data, scores on psychological tests, manifest symptoms, or other observables. The end result is a judgment; this may take the form of a recommendation concerning treatment or discharge, a decision that certain other data are necessary before final judgment is made, or a classification of the patient into a diagnostic category. What intervenes between beginning and end is, for each clinician, a quite complex idiosyncratic process. It is the purpose of this paper to demonstrate that the process is capable of rigorous investigation and description.

The Mental Process

In dealing with the manner in which clinicians utilize information at their disposal to arrive at judgments or decisions, it may appear that investigations would be concerned primarily with mental processes; and since mental processes have often been equated to or placed within the realm of subjective experience, it would be well to make one or two observations for purposes of clarification. In the first place, the term _mental process_ is often directly equated with subjective experience. But as private experience, the mental process is not observable. Hence, acceptance of this definition places the process beyond the realm of legitimate scientific inquiry, except as it may be _inferred_ from observable phenomena such as verbal responses. And since no criterion can exist for the validation of inferences concerning subjective experience, the inferences are simply ways of finding agreement in the use of language or other symbolic responses between the subject and the observer. If an observer makes "good" inferences concerning a subject, this means at most that a consensus exists between them with respect to the symbolic behavior involved. Understanding of the mental process qua subjective experience can never go beyond this level.

On the other hand, mental process may be alternately defined. It may be considered as a physical (e.g., neurological, biochemical) event capable of direct observation, i.e., using electrophysiological, neurophysiological, and similar techniques. To be sure, these techniques have so far yielded little that satisfactorily describes cognitive mental functioning, and this is perhaps unfortunate. It does not follow that the approach is sterile. Improvement in the techniques of measurement and in the application of more explanatory models may ultimately result in great progress, even though this may seem remote at the present time. This second definition is not unreasonable by any known standard, and it may surely encourage productive research and a resulting clarification of basic issues. But it is perhaps too early to say.

This brings us to the third sense in which the term mental process may be employed. It should first be pointed out that any realm of scientific investigation is designed to provide, among other things, a useful level of objective description. Direct observation, testing, instrumentation, and other related techniques are steps in this direction. When properly employed within a theoretical framework they seek to describe relationships between events or phenomena. The problem of describing judgment can similarly be considered to be one which interposes a set of techniques and a theoretical system between two sets of observables. Thus it is possible to "describe" the kinds of mental activity usually characterized as cognitive by means of mathematical models. One may thereby approach a level of description which is at least equal to that of other competitors in some respects, and certainly superior in other ways. That is to say, in controlled situations wherein the input (information) and the output (judgment) are known or capable of quantification, one may postulate functional relationships between input and output and assess their adequacy by determining the accuracy with which each is capable of predicting judgment. The present paper is directed at this level of description. The term mental process refers simply to a functional relationship which accounts for consistencies in response to divergent stimulus (information) patterns. It is thus a set of intervening variables, nothing more.

A question which immediately arises from the foregoing discussion is that of the adequacy with which it is possible to describe the mental processes underlying clinical judgment. In answer it may be said that the process is adequately described when a particular mathematical model quite effectively predicts judgments for any given set of information. This is consistent with the scientific meaning of the word "description," although considerations such as simplicity, generality, and the testability of derivations must be kept in mind. A major problem in the understanding of judgment will in this paper be considered to be that of formulating the kind of model which is sufficiently predictive, yet useful as a vehicle for approaching other related problems in the area of judgment. Different kinds of models need to be discussed, compared, and evaluated, and empirical findings are of course necessary. Subsequent sections of this paper will consider specific models for judgment. One, the linear model, is relatively simple; another to be described is somewhat more complex. Following this, some empirical findings will be offered as illustrative of the research opportunities which unfold.

Before beginning the discussion of specific models, however, it becomes necessary to justify certain restrictions that must be imposed in order that meaningful inquiry may be made into the judgment process. The restrictions center in the nature of the information available to the judge or clinician and upon which the judgment is contingent. The restrictions do not seriously impair the realism of the judgment situation as long as one can bring some ingenuity to bear upon certain problems of quantification, but even this point of view may be objectionable to some. For completeness, therefore, and in order to provide early insight into the experimental procedures used in the empirical studies to be later described, attention may now be directed to the problem of the nature of the information available to the judge.

The Information

The information upon which clinical assessment is based may commonly be expected to include anything or everything, depending upon the training and inclination of the diagnostician, and depending upon what is available or easily obtainable. The lack of control over such information may be considered an asset or a liability, depending upon one's orientation, and the accuracy of judgment may or may not be enhanced through the inclusion of non-quantitative data (Ullmann & Berkman, 1959; Holt, 1958; Luft, 1950; Meehl, 1954) depending upon the judgment domain or situation.

What seems certain, regardless of the outcomes of empirical research, is that the uncontrolled use of clinical data, whether or not it exists in quantitative form, makes clinical assessment an artistic venture. It must of course remain a matter of one's values as to whether this is wise. What seems equally certain is that any seriously conducted scientific study of judgment which has as its purpose the description of the method of combination used by the judge must take place in controlled settings, i.e., in such a way that the amount, kind, and nature of the information available to the clinician or judge can be completely specified in objective terms.

Controlling the judgment task to this degree has its advantages and its liabilitites. On the one hand, restricting the situation as described assures that each person is evaluated with respect to the same information. Ambiguous and equivocal cues are removed, and all judges are thereby certain to have at their disposal the same information and no more. The inferences made beyond this point are thus certain to have their origins in the data provided. The major problem, that of describing the idiosyncratic method of combination and weighting of this information by the clinician, is thereby clearly defined. Clinical judgments are, of course, often made in settings wherein the kinds of information available may include interview and projective test impressions, etc. In addition, the kind of information available may vary considerably from one patient to the next. This may be said to pose a limitation to the situation described. Such information may be important in judgments, a point perhaps best made by Holt (1958), but unstructured clinical judgment situations nonetheless make the contribution of such information experimentally impossible to assess. The problem here under investigation would, as a result, cease to be a scientific problem altogether.

Let us therefore consider that the situation in which a clinician makes evaluations of patients is restricted in the following ways: (a) the information available is reduced to a set of variables with respect to which all patients in the sample are evaluated; (b) the information is expressed in numbers or in categorical responses; and (c) each variable satisfies as a minimum the properties of an ordinal scale. This scheme leads to a set of numbers or classifications for each patient, where each number or class represents the degree to which a characteristic, trait, symptom, or biographical factor is present. One example of the situation satisfying these restrictions is that of a symptom check list; another is that of a test profile. Rating scale data are likewise permissible, as would be combinations of these types.

Having objectified the data upon which the judgments are to be based, we may now turn to a consideration of the model.

The Linear Model

The linear model is one in which judgments are described as a simple weighted sum of the values of the information available. For a given clinician judging a number of people, we let J represent the judgment and consider it as a dependent variable. The dimensions of information are designated by X's. These will, of course, be independent variables. If there are k sources of information, the linear additive model can be described as follows:

$$J = f(X_i) \qquad i = 1, 2, \ldots k$$

Since we are interested in a weighted sum of the X_i we may write

$$J = A_0 + A_1X_1 + A_2X_2 + \ldots + A_kX_k$$

If the A_i are so chosen as to yield the best possible weighted sum, i.e., so that the composite scores correlate maximally with J, the model is equivalent to a linear multiple regression equation wherein the weights to be applied to the independent variables are so chosen as to minimize the error in estimating an actual dependent variable from the weighted composite.

Application of multiple regression procedures to the problems of judgment has been suggested by Brunswik (1947), and by Hammond (1955). Todd (1954) reports a study using regression coefficients and the multiple correlation coefficient for a description of the clinical judgment process, where the task was to judge intelligence from a selected number of Rorschach signs. While such studies provide interesting implications, it should be stressed that there are serious limitations with respect to the interpretation of results; limitations which may be minimized or overcome only through a detailed examination of the rationale underlying the model, and through reformulations or revisions of the model, should this be necessary. So as to insure the appropriateness of the linear model as a device for characterizing the judgment process, we consider in detail some of its properties, and provide the particular reformulations where necessary.

In the first place, and by virtue of the experimental control employed in the collection of the data, the only source of reliable judgment variance is from the information supplied. This is in objective form, e.g., it appears as a number, a designated category, a position along a continuum, etc. Often these data appear as test scores on a set of protocols being judged. Assuming that a judge combined the information in linear additive fashion, the multiple regression analysis will be quite effective as a tool for describing the judgment process; i.e., the set of regression weights when applied to the corresponding predictors can quite properly serve as a model for judgment. Thus, the adequacy of the linear model can be assessed by inspection of the magnitude of multiple R. If the judge integrates data in additive fashion as opposed to configurational or pattern analysis, the linear multiple correlation will approach unity when corrected for attenuation. Lesser values of R suggest progressively lesser utility for the linear model.

Secondly, it may be noted that the regression weights signify, with certain limitations, the emphasis or importance attached to each of the predictor variables by the judge. Large coefficients mean, empirically, that the corresponding predictors can account for large proportions of the variance of judgment; and a predictor with a small beta coefficient contributes little beyond the contribution of other predictors. In practice, characterization of the judgment process by means of beta coefficients has three limitations: (a) since J's differ with respect to the size of their multiple R, direct comparisons of sets of beta coefficients between J's is not meaningful; (b) beta coefficients do not account for all the predictable variance; and (c) beta coefficients do not allow for the assessment of the independent contribution of each predictor. What would be more appropriate would be a set of weights which are comparable from one J to the next, which are capable theoretically of accounting for all of the predictable variance, and which carry exact interpretation in terms of components of variance.

Relative Weights

The formulation that is required is fortunately not difficult. Beta weights (β_{oi}) can be converted into a set of relative weights (w_{oi}) which have all the advantages described. We show first that the variance of predicted scores (which in this paper refers to predicted judgments) can be partitioned into two sources; one a sum of squared beta coefficients, and the other a residual of weighted covariances.

Let

x_o' = the predicted score for an individual (protocol) in reduced standard form.

x_i = the standard score of the ith predictor (on the protocol).

β_i = the beta coefficient for the ith predictor.

$$x'_o = \beta_{01}x_1 + \beta_{02}x_2 + \ldots + \beta_{oi}x_i + \ldots + \beta_{ok}x_k$$

or

$$x'_o = \sum_{i=1}^{k} \beta_{oi}x_i$$

$$\sigma^2_{x'_o} = \frac{1}{N}\sum^{N} [\sum_{i=1}^{k} \beta_{oi}x_i]^2$$

The term in parenthesis is a weighted variance-covariance matrix. It can be divided as follows:

$$[\sum_{i=1}^{k} \beta_{oi}x_{oi}]^2 = \sum_{i=1}^{k} \beta^2_{oi}x^2 + \sum_{i=1}^{k}\sum_{j=1}^{k} \beta_{oi}\beta_{oj}x_ix_j \qquad (i \neq j)$$

The first quantity on the right, when squared and averaged over individuals, yields the squares of the β coefficients. Thus:

$$\frac{\sum^{N}\sum_{i=1} \beta^2_{oi}x^2_i}{N} = \sum_{i=1}^{k} \beta^2_{oi}\sigma^2_{x_i} = \sum_{i=1}^{k} \beta^2_{oi}$$

since the x_i are standard scores.

Similarly, the second quantity is a weighted sum of the intercorrelations among the predictors. Thus:

$$\frac{\sum^{N}\sum_{i=1}^{k}\sum_{j=1}^{k} \beta_{oi}\beta_{oj}x_ix_j}{N} = \sum_{i=1}^{k}\sum_{j=1}^{k} \beta_{oi}\beta_{oj}r_{x_ix_j} \qquad (i \neq j)$$

Therefore,

$$\sigma^2_{x'_o} = \sum_{i=1}^{k} \beta^2_{oi} + \sum_{i=1}^{k}\sum_{j=1}^{k} \beta_{oi}\beta_{oj}r_{x_ix_j}$$

It follows that the variance of predicted scores is described by a simple sum of squared beta coefficients if and only if the covariance terms vanish. One special case in which this will be true is that of orthogonal predictors.

Relative weight, w_{oi}, is defined as follows: First we note that

$$\sqrt{\beta_{01}r_{01} + \beta_{02}r_{02} + \ldots + \beta_{ok}r_{ok}} = R_{0.12\ldots k}$$

or

$$\sqrt{\sum_{i=1}^{k} \beta_{oi} r_{oi}} = R_{0.12...k}$$

Squaring both sides and dividing by R^2, we get

$$\sum_{i=1}^{k} \frac{\beta_{oi} r_{oi}}{R^2_{0.12...k}} = 1$$

Therefore, in interpreting by independent components of variance, we express relative weight as

$$\omega_{oi} = \frac{\beta_{oi} r_{oi}}{R^2_{0.12...k}}$$

where

β_{oi} = the beta coefficient for the ith predictor.

r_{oi} = the validity coefficient (correlation with judgment) of the ith predictor

$R^2_{0.12...k}$ = the squared multiple correlation coefficient reflectint the best linear combination of the k predictors in the prediction judgments.

Finally, a description of the judgment process by means of linear regression procedures and relative weights allows one to go on to studies of varied sorts. Judges may be compared and contrasted with respect to their characteristic equations; and differences among judges may be related to training, personality, and other factors that could conceivably affect the utilization of data. Many other problems immediately suggest themselves.

The linear model may effectively be able to predict (or describe) clinical judgments to a very considerable degree, but there may be other situations for which linear models are not appropriate--just as there must be many judges for whom more complex models are necessary. Let us now turn our attention to a second type of model.

Configurational Models

In very general terms, the configurational model can be described as

$$J = f(X_1, X_2, X_3, \ldots X_k)$$

wherein the exact functional relationship involving the k independent variables may be described in any of a number of ways. As an example, let us consider a particular type of function, one which shall be referred to as an <u>interaction model.</u> The interaction model describes judgment as an appropriately weighted composite of all possible first order interactions of the predictors. Thus we may write

$$J = A_0 + \sum_{i=2}^{k} \sum_{j=1}^{k-1} A_{ij}X_iX_j \qquad (i>j)$$

The inclusion of interaction terms in a model takes account of the possibility that for a particular judge the interpretation of one item of information may be contingent upon a second. By extension, other interaction models suggest themselves. Thus it is possible to include terms involving higher order interactions or other postulated functional relationships among the information variables. Such models represent configural judgments in a quite proper sense. As the postulated interrelationships become more complex, the judgment becomes less dependent upon any single category of information and less dependent upon a simple weighted sum. Instead, the judgment comes to depend upon the configural properties of the profile, and these may approach a high degree of uniqueness. Correspondingly, the person judged is increasingly evaluated not so much with respect to a reference group of others earning the same score, but rather with respect to the <u>pattern</u> of his scores. And parenthetically, as in the case of the additive model, some of these may be judged as functionally equivalent though the scale scores differ markedly.

Other configurational models may be postulated, some of which do not involve interaction among the independent variables, but instead require a transformation of one or more of them into a different set of units. The apparent need for such transformation arises from the assumption that clinicians seldom believe that linear functions best describe relationships between information and characteristics being judged. It may, in fact, be closer to the truth that, at least for some classes of information, extreme scores are more decisive in judgment than are scores in the middle range. And clinicians will of course be expected to differ amongst themselves. For some, scores above or below some arbitrary value may carry no added significance whatsoever. In selecting configurational models for study, it therefore becomes necessary to take into account the great individual differences which may exist among clinicians and to construct that type of model which appears most promising. The application of such a model to a well-controlled situation may ultimately be capable of accounting for all but a trivial fraction of the variance of judgments.

Suppressor Effects

One of the difficulties inherent in the interpretation of beta coefficients becomes apparent when it is desired to make some statement concerning either causality or relative contribution. A beta may be high because the predictor with which it is associated correlates highly with the criterion and is relatively independent of other predictors. But it may also be high, even though the variable with which it is associated is itself a very poor predictor of the criterion, so long as its correlation with other valid predictors is sufficiently high. The point has been made in most statistics textbooks and is discussed in a few recent journal publications (e.g., Lubin, 1957). The example presented in McNemar (1955) will serve as illustrative. Assume the following:

$$r_{01} = .400$$

$$r_{02} = .000$$

$$r_{12} = .707$$

where x_0 is a judgment criterion, x_1 and x_2 are predictors (information).

Solution of this matrix leads to $\beta_{01} = .800$, $\beta_{02} = -.566$, $R_{0.12} = .566$. Quite evidently, the second predictor is a suppressor. It carries negative weight because it accounts for variance in the first predictor that is independent of the criterion. Beta coefficients would not adequately describe the judgment process for this case, since β_{02} would be -.566 even if this second predictor were unavailable as information to the judge! Indeed, it is possible, after having administered a set of protocols to a judge to add additional predictors to the correlation matrix in any arbitrary manner, and some of these might well yield significant betas.

The use of relative weights obviates this difficulty. In the example described,

$$\omega_{01} = \frac{(.400)(.800)}{(.566)^2} = 1.000$$

$$\omega_{02} = \frac{(.000)(-.566)}{(.566)^2} = .000$$

and it therefore becomes clear that a predictor must itself correlate significantly with the judgment (criterion) in order to obtain a significant relative weight.

A second point may be raised in this connection, and with reference generally to the interpretation of relative weights. Is it possible for a predictor to acquire a significant relative weight simply by virtue of its being correlated with a valid predictor? If, for example, one is asked to judge intelligence from high school rating and father's IQ (these two predictors being highly correlated), might not father's IQ turn out to have a significant relative weight by virtue of this correlation, even though the judge ignored it on the protocols?

This problem can be seen most easily in the context of partial correlation, and with reference to the extreme case. If the validity coefficient r_{01} is attributable exclusively to a second predictor,

$$r_{01.2} = 0,$$

and since

$$r_{01.2} = \frac{r_{01} - r_{02}r_{12}}{\sqrt{1 - r_{02}^2}\ \sqrt{1 - r_{12}^2}}$$

then

$$r_{01} - r_{02}r_{12} = 0$$

Further:

$$\beta_{01.2} = \frac{r_{01} - r_{02}r_{12}}{1 - r_{12}^2}$$

and:

$$\beta_{01.2} = 0$$

It would appear therefore that relative weights as defined do in fact provide meaningful descriptions of the process of judgment, i.e., with respect to the relative importance of the various items of information available to the judge, and without the kinds of spurious effects often associated with multiple regression procedures. Further, it would seem that a general mathematical proof for this conclusion is not difficult to develop. These considerations apply both to linear and configurational models.

Equivalence Among Models

A special problem becomes apparent when it is recognized that two or more models may be capable of accounting for judgment variance with equal efficiency. Consider, for example, a given model which is highly accurate in predicting judgments from the information given. In this sense we may be said to have characterized or "described" the judgmental process, but one important qualification is necessary. Even in the hypothetical situation in which prediction is perfect, one cannot conclude that the mental process has been "discovered". By definition, of course, this point should be obvious, but it is well to point out that even among sets of mathematical relationships (models) which are ostensibly different, there may be some which are in fact equivalent with respect to explanatory power.

An example may serve to clarify this point. Let us assume that for a given judge, and for two information variables, X and Y, the judgments can be independently predicted from X and Y with 95% accuracy by the following equation:

$$J' = +\sqrt{X^2+Y^2+2XY}$$

We note that the right hand term is simply the square root of the binomial $(X+Y)^2$. Since $X+Y = +\sqrt{X^2+Y^2+2XY}$, it follows that the equation $J' = X+Y$ will account for the judgments equally as well as the expression

$$J' = +\sqrt{X^2+Y^2+2XY}$$

It is therefore no more reasonable to conclude that the judge is in fact "using" one particular combination of the information than it is to conclude that he is using the other. One would have to establish different criteria before a choice between two such representations may be intelligently made.

But this should not be a troublesome point. Mathematical models are designed to provide a scheme whereby one set of events may be satisfactorily predicted from another, and whereby testable derivations may lead to more complete theoretical understanding of the phenomena. Such models therefore constitute a level of description and explanation which suffices for scientific purposes. It is not required of models that they bear any semblance of some "actual" state of affairs, either within the organism or elsewhere, nor would this necessarily lead to a better understanding of nature.

This may be more clearly seen in relation to an example from the physical sciences. A chemist has the task of describing a substance. He performs a number of operations (or tests) on the substance and determines its chemical composition. It turns out to be a relatively simple task. The substance is described, chemically, as $CaCO_3$. But the work is not complete, for a different set of operations (or tests) might have produced a different level of description; one which may not be necessary for the chemist, but which is more suitable for other contexts. The mineralogist, for example, would perform a series of tests to measure the optical properties of the crystal, or he might examine its hardness, solubility and other characteristics. Two crystals identified by the chemist as $CaCO_3$ might in fact be somewhat different from one another to the mineralogist, one being aragonite, and the other calcite. These two crystals do in fact have the same chemical structure, but they differ in molecular structure; this is revealed by optical and other tests, It is apparent that different levels of description are possible with regard to substances, and that each has its peculiar advantages and shortcomings.

In mineralogy, calcite is commonly described as a paramorph of aragonite. This word is used generally to describe a substance having crystalline structural properties which differ from those of another substance with the chemical composition.[2]

We have borrowed the term paramorphic from mineralogy and employ it in relation to representations of human judgment. The analogy may not be complete, but its limitations are not serious. The mathematical representation of the judgment process is a level of description that approaches the chemical description of minerals. The formula helps to account for or "explain" what is observed concerning certain properties or characteristics of the judge, just as the chemical formula "explains" many, though not all, properties or characteristics of the substance. In addition, the formulae are useful in making predictions concerning the outcomes of certain other tests which may later be employed. But as with chemical analysis, the mathematical description of judgment is inevitably incomplete, for there are other properties of judgment still undescribed, and it is not known how completely or how accurately the underlying process has been represented. The term "paramorphic representation," used in relation to judgment, would seem adequately to indicate this state of affairs.

The Linear Model: Representative Results

As illustrative of the significance of the methodology described above, four persons who participated as judges in studies at Oregon have been selected.

Two of these made judgments of "intelligence" of 100 persons on the basis of a set of nine predictors or sources of information. The remaining two made judgments of "sociability" of 150 persons on the basis of profiles containing scores on eight selected Edwards Personal Preference Schedule (EPPS) variables. In all cases, the judges returned after an intervening period of several days and made a second set of judgments on the same profiles. Sample profiles are shown in Figure 1 and Figure 2.

Insert Figures 1 and 2 about here

Application of standard multiple regression procedures reveals that a best linear combination of the predictor scores correlates .948 and .829 with the judgments of Judges 15 and 18, respectively.[3] But to what extent can it be said that the linear model adequately characterizes the judgment process for these judges? The reliability of judgment for these two is .876 and .836 respectively. Correction for attenuation results in coefficients which are 1.00+ and .907. Thus it may be seen, in the case of the first judge, that a linear additive combination of predictors (once a "best" set of regression weights is known) allows the prediction of judgment with virtually complete certainty. The residual or error variance is trivial in comparison with that which is predictable from the model. For the second judge the case is somewhat different. When unreliability of judgment is taken into account, there still remains 17.7% of variance which is unaccounted for by the model. It follows, therefore, that the linear model, while quite sufficient for the first judge, is less appropriate for the second.

A second type of question may be asked with respect to these judges: Is a judge able to describe with any ascertainable degree of validity the manner in which he utilizes information in arriving at his judgment? In its present form this question is probably unanswerable, but it is quite proper to ask a related one. Concerning the method of utilization of information, what correspondence exists between the verbal description offered by the judge and the description achieved by the multiple regression model?

There are some difficulties in attempting to ascertain, from the statements of the judge, a subjective impression of a cognitive process. In instances in which persons are asked to make judgments of intelligence and sociability from the sets of information that have just previously been described, it is rarely true that the judge has a high degree of confidence in statements he may make concerning the relative importance of the predictor variables. In some instances difficulties of communication emerge; when, for example, a judge finds it necessary to relate some rather complex or configurational analysis that he feels best describes his own "method of combination." One alternative is to ask the judge to distribute 100 points among the sources of information available and in such a way that this distribution reflects, to the best of his knowledge, the relative importance of those variables.

Such a task is easily understood by the judge. The method has the additional advantage of insuring that his subjective impressions are personally translated into numerical form without interpretation by a second person. Presumably, some information is lost in the process, as would be the case wherein a judge does not consider his method of combination to be adequately describable by a weighted sum of the information. Notwithstanding such attitudes, however, subjects used as judges in studies presently underway commonly report satisfaction and confidence in the procedure, and without apparent relationship to their attitudes concerning the complexity of the method of combination they believe themselves to be using.

The number of points assigned in this way to each of the information variables will be referred to as the subjective weight (s_{oi}). Comparisons of subjective and relative weights are shown for the two judges of intelligence in Figures 3 and 4.

Insert Figures 3 and 4 about here

With respect to Judge 18, there is a high degree of agreement of relative and subjective weights. For Judge 15, however, there are greater discrepancies, disagreement being most pronounced in Variables 1, 4, and 7. These judges differ in the extent to which they are capable of assigning a set of numbers to the sources of information used in judgment so as to approximate the relative weights as determined by the linear model. Results from a relatively large sample of judges will be reported in a forthcoming article.

The two examples from the sociability experiment can be used to illustrate the same phenomena. The relevant information is described in Table 1 and in Figures 5 and 6.

Insert Table 1 and Figures 5 and 6 about here

A Configurational Model: Representative Results

Data for one type of configurational model comes from a study by Martin (1957). On the basis of lengthy and rather complete interviews, Martin was able to obtain fairly clear statements from a set of five counselling psychologists. These statements expressed the manner in which the psychologists believed they were utilizing the set of eight EPPS variables in the prediction of sociability. Of the five judges, Psychologist D has been selected for illustrative purposes. His verbalizations with respect to the judgment of sociability follow:

> As might be expected I generally look for some over-all patterning of the test variables. Although I do hold in mind certain standards and/or tendencies. In the following discussion when I refer to high I mean at least 1 SD above the mean. For a rating of high Sociability, i.e., 7, 8, or 9, I would generally expect at least two or three scales (Exh. + Dom. + Het.) to meet the criterion of 1 SD above the mean with others at least in the middle range or pointing in the direction of high. I would also expect for this rating that Aba. be in the average range or in the direction of high. I would also expect for this rating that Aba. be in the average range or in the direction of low. The more that Exh. + Dom. + Het. approach the high extreme and Aba. the low extreme the more apt I

am to make a higher rating. In this concept the other four scales act in pairs. For example, if the above conditions are met and Def. and Suc. are not in either extreme the rating remains unaffected. If both are extremely low they tend to add to the rating and if both are extremely high they tend to detract from the rating of high Sociability.

The scales Chg. and Aff. are also considered as a pair but add or detract nothing to the ratings unless both are quite high or quite low. If both should be extremely low and the other conditions for a high rating are met I would suspect the reliability of the test.

In the case of a low rating (1, 2, 3) I would generally expect a somewhat opposite patterning. For example, here I would expect Aba. to be quite high with pair Suc. and Def. in the middle range or pointing in the direction of high. Similarly, I would expect at least two of the three scores (Dom., Exh., and Het.) to be in the middle or low range with none of them extremely high. The higher Aba. along with the pair Def. and Suc. and the lower the scales Dom., Het., and Exh., the lower the rating. The pair Chg. and Aff. again have little effect unless they are significantly low and then they tend to support or add to a low rating.

In rating within the average range I look for the significant scales Exh., Dom., Het., and Aba., not to be extreme in either direction. While the variables are considered in relation to one another (a high score on Exh. offsets a high score on Aba.) they contribute to my final rating whether singly or in pairs.

From the foregoing description, it may be argued that the set of variables underlying D's judgments must take into account the following considerations:

1. Interaction. Certain of D's statements imply an interactive or multiplicative relationship between two predictors and the judgment criterion. A good example of this is the statement, "The scales Chg. and Aff. are also considered as a pair but add or detract nothing to the rating un-

less both are quite high or quite low."

2. Nonlinearity. Since D has stated that variables are of most importance when their values are high, and that values between $\pm$ 1σ of the mean are usually ignored, an exponential function should be more predictive of judgment than a linear one.

As a result, the following predictors are defined:

X_1 Abasement

X_2 Exhibitionism

X_3 Heterosexuality

X_4 Dominance

X_5 Change x Affiliation

X_6 Succorance x Deference

Each of the six predictors can then be subjected to a nonlinear transformation, the result of which is to correct for the stated tendency of the judge to discount predictor scores near the mean of the distribution, and to emphasize them increasingly as they become extreme, up to a limit. The resulting variables are then weighted according to least squares procedures.

As in the case of the linear model, it is possible to compute relative weights and to compare these with the weights assigned subjectively. This comparison is shown in Figure 7. We shall discuss the data shortly.

Insert Figure 7 about here

What classes of problems can be attacked through the technique of the configurational model? As in the case of the linear model, it furnishes a description of the relative importance (to the judge) of the various sources of information available. But greater latitude is possible. Configurational models are capable of handling the complexities and patterns believed by many to be an essential (if not "natural") part of the judgment process. Thus, one may ask whether a graduate course in psychodiagnostics or in personality assessment was effective in producing students who are "configurational" in their interpretation of case material.

A second class of problem is that of individual differences. Do persons differ in the type of model which most appropriately accounts for their judgments? If so, in what respects? And what proportion of these differences can be attributed to specific training? To personality? To intellectual characteristics?

There is a third class of problem, one which is at least of equal significance to the others mentioned. This concerns the stability of the judgment process. Can a "linear" judge be taught to be more complex? Can any judge be taught to make more efficient and more accurate use of information? Will this generalize or transfer to other judgment situations? What are the personality characteristics which differentiate the flexible judge from the unchangeable one? Some of these problems are currently under investigation and results will appear in forthcoming publications.

Finally, it may already have become apparent that the issue of nomothetic vs. idiographic approaches, first proposed by Windelband (1904) and discussed more thoroughly elsewhere (Allport, 1937; Sarbin, 1944) may be approached in one of its major aspects through the use of judgment models. If the argument is confined to that of "method of combination," as suggested by Meehl (1954), representational models are capable of providing some strikingly clear evidence. Martin (1957) presents some interesting findings in this respect, and it may be well for illustrative purposes to return to Psychologist D, discussed above, and ask what the data suggest.

The first thing that might be said with respect to this psychologist is that, after having defined the predictors to his liking, he is not as astute as he might be in attaching subjective weights to them. By this it is meant that the discrepancies between subjective and relative weights are by no means insignificant in an absolute sense. The matter can be made even clearer by stating it another way:

Given the variables in question, a computing machine would come closer to reproducing D's judgments than he could come himself, were he to do the weighting as he says it should be done. What may be of particular importance is the fact that both of the interaction variables were overevaluated in this respect. Perhaps one likes to think of himself as being more complex than he actually is.

Since the computer does as well as it does in reproducing the judgments, an example is here provided wherein a single consistent set of rules of combination (nomothetic) may be successfully applied to the many and varied cases. The resulting judgments may well differ from "configural" or "clinical" judgments by amounts which are quite trivial. It is not impossible that this phenomenon will occur even in instances wherein the judge is schooled in the idiographic tradition and believes himself to be functioning accordingly. But this generalization and the earlier ones are perhaps reckless at this point, particularly in that the error variance for relative and subjective weights is not presented for these data. The purpose of the paper is more to illustrate the application of a methodology than to argue for reforms in clinical psychology from evidence based on a sample of Size 1.

What, additionally, may be said of the configurational model developed for D? We may legitimately ask to what extent the substitution of complex variables in the place of simple ones effectively enhanced the prediction of judgments, thereby testing the relative efficacy of two models. The R (corrected for shrinkage), using the configurational model, is .88. This appears high, and it may well be, except for the fact that the application of a linear model to D's judgments results in a corrected R of .91. Thus, substituting a configurational model for a linear one changes the proportion of predicted variance of judgments from 82.81% to 77.44%; a loss of a little better than 5%! And considering chance factors, the least that can be said is that there is no demonstrable gain over the linear model.

Summary

This paper has been concerned with the manner in which information is utilized in decision making or in judgment situations. It is shown that mathematical models provide a way of describing mental processes which would otherwise be accessible only through introspection or electro-physiological techniques. A linear model and a configurational model are described, and illustrations furnished for each. Such models make possible the testing of hypotheses concerning method of combination, individual differences in judgment ability, effects of training, personality correlates, idiographic interpretation of case materials, etc.

It may be said that the paramorphic representation of judgment appears to offer interesting possibilities for research and theory as well as for applied endeavors. Models which attempt to describe the method of combination of information in decision making can illuminate problems which would otherwise remain obscure. In focusing upon the individual as the unit of research while at the same time preserving methodological rigor it becomes possible to achieve a level of psychological description which would otherwise be quite difficult. And few would disagree with the suggestion that sound description of the decision process is quite fundamental to a complete understanding of man.

Footnotes

1. This investigation was supported by a Public Health Service research grant (M2097-C1) from the National Institute of Mental Health.

The analysis of much of the data referred to in this report was made possible through the facilitites of the Division of Counseling & Testing Services, University of Washington, and the Western Data Processing Center at the University of California, Los Angeles.

2. More exactly, the crystal is called a paramorph when it is shown to be an alteration in crystal structure. In the example cited, aragonite may undergo change over a period of time, finally becoming calcite. It is the calcite that results from alteration of aragonite which is paramorphic.

3. Corrected for shrinkage. Multiple R's computed from a cross-validation sample of protocols were .937 and .837 for these J's respectively.

References

Allport, G. W. _Personality, a psychological interpretation._ Holt, 1937.

Brunswick, E. _Systematic and representative design of psychological experiments._ Berkeley: University of California Press, 1947. (Also in J. Neyman (Ed.), _Berkeley symposium on mathematical statistics and probability_. Berkeley: University of California Press, 1949. Pp. 143-202.)

Hammond, K. R. Probabilistic functioning and the clinical method. _Psychological Review_, 1955, _62_, 255-262.

Holt, R. R. Clinical and statistical prediction: A reformulation and some new data. _Journal of Abnormal and Social Psychology_, 1958, _56_, 1-12.

Lubin, A. Some formulae for use with suppressor variables. _Educational Psychological Measurement_, 1957, _17_, 286-296.

Luft, J. Implicit hypotheses and clinical predictions. _Journal of Abnormal and Social Psychology_, 1950, _45_, 756-759.

McNemar, Q. _Psychological statistics._ Wiley, 1955.

Martin, H. T., Jr. The nature of clinical judgment. Unpublished doctoral dissertation, Washington State College, 1957.

Meehl, P. E. _Clinical vs. statistical prediction._ Minneapolis: University of Minnesota Press, 1954.

Sarbin, T. R. The logic of prediction in psychology. _Psychological Review_, 1944, _51_, 210-228.

Todd, F. J. A methodological study of clinical judgment. Unpublished doctoral dissertation, University of Colorado, 1954.

Ullmann, L. P., & Berkman, Virginia C. Judgments of outcome of home care placement from psychological material. _Journal of Clinical Psychology_, 1959, _15_, 28-31.

Windelband, W. _Geschichte und Naturwissenschaft._ (3rd ed.) 1904.