ORGANIZATIONAL BEHAVIOR AND HUMAN PERFORMANCE 6, 458–479 (1971)

Five Models of Clinical Judgment: An Empirical Comparison between Linear and Nonlinear Representations of the Human Inference Process[1]

LEWIS R. GOLDBERG
University of Oregon and Oregon Research Institute, Eugene, Oregon 97403

Two nonlinear models, which were proposed by Einhorn as approximations of the Conjunctive and Disjunctive strategies, were compared with the Linear, Logarithmic, and Exponential models as potential representations of the judgments made by 29 clinical psychologists, each of whom attempted to differentiate neurotic from psychotic patients on the basis of their MMPI profiles. For this task, the Linear model provided a better representation of the judgments made by all clinicians than did either of Einhorn's models, and only the Logarithmic provided the Linear model with any real competition. Moreover, serious problems endemic to all of the nonlinear models call into question their utility as potential judgmental representations.

How do physicians and clinical psychologists integrate the data they receive from each of their patients to make their diagnostic decisions? What policies are used by college admissions committees and personnel officers to arrive at selection decisions? In what manner does the average individual combine data in making job and family decisions? Stimulated by the seminal contributions of Hammond (1955) and Hoffman (1960), the past decade has witnessed an explosion of research interest in answering such questions (e.g., Anderson, 1962, 1965, 1967–1969; Hammond, Hursch & Todd, 1964; Hammond & Summers, 1965; Hoffman, 1968; Hoffman, Slovic & Rorer, 1968; Naylor, 1964, 1967; Naylor & Schenck, 1968; Naylor & Wherry, 1965; Slovic, 1966, 1969; Tucker, 1964; Wiggins & Hoffman, 1968).

Over the years, various mathematical models have been proposed as potential representations of the judgment process. Empirical comparisons

[1] This study was supported by Grant No. MH 12972 and MH 10822 from the National Institute of Mental Health, U. S. Public Health Service. Computing assistance was obtained from the Health Sciences Computing Facility, UCLA, sponsored by NIH Special Resources Grant RR-3.

among competing models have yielded one rather consistent finding: across a fairly broad array of judgmental tasks, a linear and compensatory model (multiple regression analysis) has consistently provided at least as faithful a representation of human inferences as have other models tried to date (Goldberg, 1968). However, since it is intuitively unreasonable to assume that man typically processes information in the same manner as a linear regression analysis (Hoffman, 1968), *some* type of nonlinear and/or noncompensatory models should eventually prove superior to the linear model—at least for some judges with some sorts of inferential tasks. Consequently, two recent papers by Einhorn (1970, 1971) are of unusual importance, since they purport to have demonstrated that for a sizable proportion of judges a nonlinear model provided a more accurate judgmental representation than did the linear regression model.

Einhorn (1971) compared two nonlinear models (Conjunctive and Disjunctive) with the Linear model as representations of the judgments made by (*a*) each of 39 engineering students, who rank ordered 15 jobs in terms of their attractiveness, and (*b*) each of 30 Ph.D.-level psychologists, who rank ordered 15 hypothetical applicants to graduate school in terms of their likely acadamic promise. Equal numbers of judges were assigned to each of three treatments, in which they received either two, four, or six cues to use in making their rankings. Each judge ranked two different sets of 15 protocols, and the three models were compared in a double cross-validation design. Einhorn concluded from this study that (*a*) the average accuracy of each of the two nonlinear models relative to that of the linear model did not change as a function of the number of cues provided the judges, (*b*) all three models provided less accurate judgmental representations as the number of cues increased,[2] and (*c*) one of the two nonlinear models provided a better representation of the judgments made by many judges than did the linear model, a finding which was most striking in the job preference study.

However, Einhorn constructed his cue sets specifically to reveal any discrepancies between linear and nonlinear judgmental strategies. As a result, cues which have high correlations in nature (e.g., Verbal and Quantitative scores from the Graduate Record Examination) may have

[2] This conclusion was based upon the finding that the size of the correlations between the models and the judgments decreased as the number of cues increased. This finding does not necessarily imply that the models provided any poorer representations of the reliable variance in the judgments as the number of cues increased, since the judges simply may have been less reliable in the six-cue than in the two-cue task. Unfortunately, since Einhorn did not collect test-retest reliability judgments, this hypothesis must remain untested.

been negatively correlated within the set of profiles Einhorn presented to his judges. Consequently, it is difficult to generalize the findings from these small sets of unusual cases to judgments made from a larger and more representative set of protocols. The present study was designed to extend Einhorn's findings to a much larger and more representative set of data. Moreover, the present study permits a comparison among the three models used by Einhorn, plus two "control" models, which were not included in Einhorn's study. In addition, the present study is focused upon a judgmental problem of considerable societal significance (the differential diagnosis of psychosis vs neurosis), and, therefore, this task provides an important arena for discovering the inferential strategies of experienced diagnosticians.

Specifically, the present study was designed to investigate which of five models (Linear, Conjunctive, Disjunctive, Logarithmic, and Exponential) provides the most accurate representations of the differential diagnoses made by each of 29 clinical psychologists, based on the MMPI profiles of 861 psychiatric patients. While previous research using the same data (e.g., Goldberg, 1965, 1969) has shown that the linear model is not surpassed by any of a host of nonlinear prediction schemes in forecasting the *criterion* diagnoses, only a few studies (e.g., Wiggins & Hoffman, 1968) have focused upon the 29 sets of clinical judgments themselves. Wiggins and Hoffman (1968) compared the linear regression model with two competitors (*a*) a quadratic regression function, and (*b*) a regression function composed of "patterns" and "signs" culled from the psychometric literature. Since these investigators concluded that "configurality was a consistent judgmental characteristic distinguishing 16 of the judges" (Wiggins & Hoffman, 1968, p. 70), these data should be particularly appropriate as a test of the nonlinear models proposed by Einhorn.

METHOD

The Judgmental Task

The data for this study were collected by Paul Meehl, who originally focused research attention on this problem on the grounds that "the differences between psychotic and neurotic profiles are considered in MMPI lore to be highly configural in character, so that an atomistic treatment by combining scales linearly should theoretically be a very poor substitute for a configural approach" (Meehl, 1959, p. 104). Meehl collected 861 MMPI profiles from seven hospitals and clinics throughout the country, each profile from a psychiatric patient who had been diagnosed as being either psychotic or neurotic. For a description of the

seven samples, see Goldberg (1965). Of the 861 MMPI profiles, approximately half (47%) were produced by psychotic patients.

The Judges

Meehl obtained diagnostic judgments from 29 clinical psychologists (13 at the Ph.D. level and 16 advanced graduate students at the University of Minnesota). The judges rated the profiles from each sample in turn, using an 11-step forced-normal distribution, from least (likely) to most (likely) psychotic. A composite judge was constructed for the present study by averaging the ratings of the 29 clinicians for each profile.

The Judgmental Models

The five models which were compared in this study are presented in Table 1. Both the conceptual and the computing formulas are listed in the table; for two of the five models (the Linear and the Logarithmic) the conceptual and computing formulas are identical. The first three models listed in Table 1—the Linear, Conjunctive, and Disjunctive

TABLE 1
THE FIVE MODELS[a]

Name	Conceptual formula	Computing formula
Linear (LIN)	$\hat{Y} = \sum_{i=1}^{k} b_i X_i$	$\hat{Y} = \sum_{i=1}^{k} b_i X_i$
Conjunctive (CON)	$\hat{Y} = \prod_{i=1}^{k} X_i^{b_i}$	$\log \hat{Y} = \sum_{i=1}^{k} b_i \log X_i$
Disjunctive (DIS)	$\hat{Y} = \prod_{i=1}^{k} \left(\frac{1}{a_i - X_i}\right)^{b_i}$	$\log \hat{Y} = -\sum_{i=1}^{k} b_i \log (a_i - X_i)$
Logarithmic (LOG)	$\hat{Y} = \sum_{i=1}^{k} b_i \log X_i$	$\hat{Y} = \sum_{i=1}^{k} b_i \log X_i$
Exponential (EXP)	$\hat{Y} = \prod_{i=1}^{k} e^{b_i X_i}$	$\log \hat{Y} = \sum_{i=1}^{k} b_i X_i$

[a] $\hat{Y}$ is the predicted judgment for each judge and each profile; X_i is the value of the i-th cue (in this case, an MMPI scale score); b_i is the regression weight for the i-th cue; a_i is an arbitrary constant (in this case, a number one scale point larger than the highest obtained value of the i-th cue); k is the number of cues (in this case, 11); and all scales have been reflected, if necessary, so that $r_{XY} \geq 0$.

models—were taken directly from Einhorn (1970, 1971). The two remaining models—the Logarithmic and Exponential—were included as "controls" for the Conjunctive model, to check whether any possible incremental validity of the Conjunctive over the Linear model might simply stem from the logarithmic transformation of the cues alone (LOG) or of the judgments alone (EXP). The two parallel "controls" for the Disjunctive model are not included, since neither Einhorn's research, nor that to be reported here, showed any substantial incremental validity for that model.

The Design of the Study

For each judge in turn, multiple correlations were obtained between his actual judgments and the predicted values from each of the five models, both within each of the seven samples of MMPI profiles and for the total sample of 861 profiles. In addition, the same values were obtained for the average judge, the composite judge, and the criterion diagnoses. Since these multiple correlations were all based upon the same sample of cases used to derive the models, the values will necessarily be higher than they would be in a cross-validation sample. However, since all five models include the same number of parameters (11), their shrinkage upon cross-validation should be relatively uniform. Nonetheless, as a further test of the comparative utility of the models, the 861 profiles were split into two parallel sets of 430 and 431 profiles, and the same analyses were carried out using a double cross-validation design. That is, the models derived in one set of profiles were cross-validated in the other set (and vice versa), and the two cross-validated correlations were then averaged.[3]

RESULTS

Table 2 presents the multiple correlations for each of the five models, based upon the total sample of 861 cases. Values are presented for each of the 29 judges, the average judge, the composite judge, and the criterion diagnoses. Note that for all 29 judges (100%), the Linear model provided a better representation of their judgments than did either the Conjunctive or the Disjunctive models. In fact, for 25 of the 29 judges (86%), the Linear model provided a better fit than *any* of the other four models. For four judges, the Logarithmic model was slightly superior to the Linear one, though all of these minute differences appear only in the third decimal place. In general, for both the average and the composite

[3] All computations were carried out by computer in double-precision arithmetic, a practice suggested for all investigators using logarithmic functions.

TABLE 2
A Comparison of the Multiple Correlations for Each of Five Models
Total Sample (N = 861)

Judge	LIN	CON	DIS	LOG	EXP
1	.70*	.63	.60	.68	.65
2	.83*	.74	.64	.82	.75
3	.72*	.62	.55	.71	.63
4	.72*	.63	.62	.68	.67
5	.77*	.69	.60	.76	.69
6	.80*	.71	.69	.77	.74
7	.75*	.67	.60	.74	.67
8	.81	.75	.61	.81*	.73
9	.82*	.75	.67	.81	.74
10	.73	.72	.62	.73*	.71
11	.84	.79	.68	.84*	.78
12	.76*	.65	.60	.74	.66
13	.74*	.63	.57	.72	.64
14	.58*	.44	.52	.54	.49
15	.83*	.75	.68	.80	.78
16	.84*	.76	.69	.83	.76
17	.75*	.62	.63	.67	.70
18	.86*	.78	.63	.84	.77
19	.79*	.73	.63	.78	.72
20	.78*	.72	.64	.77	.72
21	.85*	.77	.70	.84	.77
22	.80*	.74	.67	.79	.73
23	.75*	.63	.55	.72	.66
24	.70*	.62	.54	.69	.63
25	.84	.79	.67	.84*	.77
26	.78*	.70	.56	.77	.70
27	.74*	.66	.61	.73	.66
28	.82*	.75	.63	.81	.75
29	.78*	.74	.63	.78	.73
Average	.78*	.70	.62	.76	.70
Composite	.89*	.86	.75	.88	.86
Criterion	.46*	.40	.41	.40	.46

* Best-fitting model.

judge, the Linear and the Logarithmic models provided the best fit, while the Disjunctive model turned out to provide the worst representations of any of the five models.

Table 3 provides the parallel values based upon the average cross-validated correlations. The values in Table 3, as expected when the number of cases is this large, are virtually identical to those presented in

TABLE 3
AVERAGE CROSS-VALIDITY COEFFICIENTS FOR EACH OF FIVE MODELS

Judge	LIN	CON	DIS	LOG	EXP
1	.69*	.61	.56	.66	.64
2	.82*	.73	.56	.81	.74
3	.71*	.62	.49	.70	.63
4	.72*	.61	.56	.67	.66
5	.76*	.68	.52	.76	.68
6	.80*	.69	.62	.76	.73
7	.74*	.65	.56	.72	.66
8	.80	.74	.53	.81*	.72
9	.81*	.74	.62	.81	.73
10	.72	.70	.54	.72*	.69
11	.84*	.78	.61	.84	.77
12	.75*	.65	.53	.74	.66
13	.72*	.61	.49	.71	.61
14	.55*	.40	.49	.51	.45
15	.83*	.73	.60	.78	.77
16	.84*	.75	.64	.82	.76
17	.74*	.62	.55	.68	.68
18	.85*	.75	.56	.82	.75
19	.78*	.72	.57	.77	.71
20	.77*	.71	.57	.76	.70
21	.84	.76	.64	.83	.76
22	.79*	.71	.62	.78	.72
23	.74*	.61	.49	.70	.64
24	.69*	.61	.46	.67	.61
25	.83*	.78	.61	.83	.76
26	.76*	.68	.46	.75	.67
27	.73*	.62	.55	.71	.64
28	.81*	.74	.51	.81	.74
29	.77*	.73	.55	.77	.73
Average	.77*	.68	.55	.75	.69
Composite	.88*	.86	.68	.87	.86
Criterion	.41*	.34	.36	.34	.41

* Best-fitting model.

Table 2. The average shrinkage for the judgments was about .01, and for the criterion about .05. Again, the Linear model provided a better fitting representation than the Conjunctive or Disjunctive models for all 29 judges, and for 27 of the 29 judges (93%) the Linear model provided the best fit of all five models. Even for the two judges who were better represented by the Logarithmic model, once again the differences were minute.

TABLE 4
THE BEST-FITTING MODEL FOR EACH JUDGE IN EACH SAMPLE

	Sample							
Judge (N)	A (92)	B (77)	C (103)	D (42)	E (181)	F (166)	G (200)	Total (861)
1	LIN	LOG	LIN	CON	LIN	CON	LIN	LIN
2	LOG	LIN	LIN	CON	LIN	LIN	LIN	LIN
3	LOG	LIN	LIN	LIN	LIN	LIN	LIN	LIN
4	LIN	LIN	LIN	LIN	LIN	LIN	LIN	LIN
5	LOG	LIN	LIN	LIN	LIN	LIN	LIN	LIN
6	LOG	LIN	LIN	CON	LIN	LIN	LIN	LIN
7	LIN	LOG	LIN	CON	LIN	LIN	LIN	LIN
8	LOG	LOG	LOG	LIN	LIN	LIN	LIN	LOG
9	LIN	LOG	LOG	LIN	LIN	LIN	LIN	LIN
10	LOG	LOG	LOG	DIS	LIN	LIN	LOG	LOG
11	LOG	LOG	LOG	LOG	LIN	LIN	LIN	LOG
12	LOG	LIN	LIN	LIN	LIN	LIN	LIN	LIN
13	LIN	LIN	LIN	LIN	LIN	LIN	LIN	LIN
14	LIN	DIS	LIN	CON	DIS	LIN	LIN	LIN
15	LIN	LIN	LIN	CON	LIN	LIN	LIN	LIN
16	LIN	LIN	LIN	LOG	LIN	LIN	LIN	LIN
17	LIN	LIN	LIN	LIN	LIN	LIN	LIN	LIN
18	LOG	LOG	LIN	LIN	LIN	LIN	LIN	LIN
19	LOG	LOG	LOG	LIN	LIN	LIN	LIN	LIN
20	LOG	LOG	LIN	LIN	LIN	LIN	LIN	LIN
21	LIN	LOG	LIN	LIN	LIN	LIN	LIN	LIN
22	LOG	LOG	LOG	LIN	LIN	LIN	LIN	LIN
23	LOG	LOG	LIN	LIN	LIN	LIN	LIN	LIN
24	LOG	LIN	LOG	LIN	LIN	LIN	LIN	LIN
25	LOG	LOG	LOG	LOG	LIN	LIN	LIN	LOG
26	LOG	LOG	LOG	LIN	LIN	LIN	LIN	LIN
27	LIN	LIN	LIN	LIN	LIN	LIN	LIN	LIN
28	CON	LIN	LIN	LIN	LIN	LIN	LIN	LIN
29	LOG	LIN	LOG	LIN	LIN	LIN	LIN	LIN
Average	LOG	LIN	LIN	LIN	LIN	LIN	LIN	LIN
Composite	LOG	LIN	LOG	LIN	LIN	LIN	LIN	LIN
Criterion	LOG	LIN	LIN	LIN	LIN	LIN	LIN	LIN

Parallel tables based upon each of the seven samples of MMPI profiles are available from the author,[4] and a summary of these results is presented in Table 4. For each of the seven samples, and for the total sam-

[4] These additional tables may be obtained without charge from the author at Oregon Research Institute (P. O. Box 3196, Eugene, Oregon 97403) or for a fee from the American Documentation Institute.

ple, the model which provided the best representation for each judge is listed. By reading down the columns of Table 4, the reader can locate any inter-sample differences. For the four smallest samples (A, B, C, and D), 54% of the comparisons favored the Linear model, 38% the Logarithmic model, 6% the Conjunctive model, 2% the Disjunctive model, and 0% the Exponential model. However, for the three larger samples (E, F, and G), 97% of the comparisons favored the Linear model. By reading across the rows of Table 4, the reader can locate any inter-judge differences. For example, four judges (4, 13, 17, and 27) were best represented by the Linear model in all seven samples, while most judges were represented by the Logarithmic model in a few samples and the Linear model in the rest. While no judge was best represented by either the Conjunctive or Disjunctive models in more than two samples, three judges (10, 11, and 25) were best represented by the Logarithmic model in five of the seven samples. In general, however, inter-sample differences were more pronounced than inter-judge differences.

In all of the above analyses, the 11 MMPI scale scores (the cues) were each expressed in T-score form, with values ranging from about 40 to 120. Since any linear change of scale will affect the accuracy of the Conjunctive, Disjunctive, and Logarithmic models, a final series of analyses was carried out after linearly rescaling each of the cues so that its lowest value was 1. The results from one such analysis are presented in Table 5. This table, which can be compared directly with Table 2, presents the multiple correlations for each of the five models based upon the total sample of 861 profiles. Note that when the cue values were rescaled, all of the 29 clinicians were better represented by the Linear model than by *any* of the nonlinear ones. The multiple correlations based on the Linear and Exponential models were, of course, not changed by cue rescaling. On the other hand, the correlations for the Conjunctive and Logarithmic models were somewhat lower, and those for the Disjunctive somewhat higher, after the cues were rescaled (Table 5), than prior to rescaling (Table 2).

As has been argued elsewhere (e.g., Goldberg, 1968; Hoffman, 1960, 1968), the power of the Linear model in representing these judgments must not be construed as implying that the judges were actually processing the cues in a linear and compensatory fashion. In fact, there is additional evidence indicating that at least some of the variance in the judgments of a number of these clinicians was clearly nonlinear in character, a finding which makes the present "victory" of the Linear model all the more significant. Specifically, Wiggins and Hoffman (1968) concluded that sets of linear and nonlinear signs culled from the MMPI literature provided better judgmental representations for some of the

TABLE 5
A Comparison of the Multiple Correlations with the Cues Rescaled
Total Sample (N = 861)

Judge	LIN	CON	DIS	LOG	EXP
1	.70*	.61	.64	.65	.65
2	.83*	.70	.72	.77	.75
3	.72*	.58	.62	.66	.63
4	.72*	.58	.67	.63	.67
5	.77*	.65	.67	.71	.69
6	.80*	.66	.74	.71	.74
7	.75*	.63	.66	.69	.67
8	.81*	.73	.70	.77	.73
9	.82*	.73	.72	.78	.74
10	.73*	.71	.68	.71	.71
11	.84*	.78	.75	.81	.78
12	.76*	.62	.65	.69	.66
13	.74*	.59	.62	.68	.64
14	.58*	.40	.52	.48	.49
15	.83*	.72	.75	.75	.78
16	.84*	.72	.75	.78	.76
17	.75*	.57	.69	.62	.70
18	.86*	.74	.73	.79	.77
19	.79*	.71	.70	.75	.72
20	.78*	.68	.70	.71	.72
21	.85*	.74	.75	.80	.77
22	.80*	.71	.72	.76	.73
23	.75*	.56	.65	.65	.66
24	.70*	.59	.61	.64	.63
25	.84*	.78	.74	.81	.77
26	.78*	.64	.66	.71	.70
27	.74*	.64	.65	.70	.66
28	.82*	.72	.72	.77	.75
29	.78*	.70	.71	.74	.73
Average	.78*	.66	.68	.71	.70
Composite	.89*	.82	.83	.83	.86
Criterion	.46*	.37	.46	.37	.46

* Best-fitting model.

judges than did the Linear model alone. As a further test of this finding, the judgments of each of the 29 psychologists, plus the composite judge and the criterion diagnoses, were each converted to residual scores, the linearly predictable variance from the 11 MMPI scale scores having been partialled out of the residuals. These residual scores thus contain that variation in the original judgments which is not predictable from a linear combination of the 11 cues. When each of these 31 residual scores

were correlated with the MMPI signs described in Goldberg (1965), a number of highly significant correlations were discovered. The significant findings from two of these analyses, those focused upon the residual composite judge and the residual criterion diagnoses, are presented in Table 6. Note that for the residual composite judge (the judgmental consensus of all 29 clinicians) 13 nonlinear signs produced significant correlations, although only the simplest of these signs—the High Point Rules (if the highest clinical scale on the profile is either *Hs, D, Hy,* or *Pt,* diagnose Neurotic; if not, diagnose Psychotic)—had any significant correlation with the residual criterion diagnoses.

TABLE 6
SIGNIFICANT CORRELATIONS BETWEEN MMPI NONLINEAR SIGNS AND (a) THE RESIDUAL COMPOSITE JUDGE AND (b) THE RESIDUAL CRITERION DIAGNOSIS ($N = 861$)

Sign	Residual judge	Residual criterion
High Point Rules	.23**	.08*
No. of clinical scales < 45	.23**	.04
Variance of *Pa, Pt,* & *Sc*	.22**	.03
Peterson signs	.21**	.05
Rank of *Ma*	−.12**	−.02
Taulbee-Sisson signs	−.10**	−.03
No. of clinical scales ≥ 55	−.09*	−.04
$F \times Sc$	−.09*	.02
$Pd \times Sc$	−.09*	.01
$K \times Sc$	−.08*	.01
$L \times Sc$	−.08*	.03
Rank *Pa*	−.08*	−.02
Rank *D*	.08*	.03

Note: For an explanation of each of the signs, see Goldberg (1965).
* $p < .05$.
** $p < .01$.

DISCUSSION

The findings from the present study seem reasonably clear: The Linear model generally provided a better representation of the diagnostic judgments of these 29 clinical psychologists than did either the Conjunctive or the Disjunctive models. And, of the five models utilized in this study, only the Logarithmic model provided the Linear model with any real competition.

Seemingly, these findings are in direct contradiction to those of Einhorn (1971). Since the case for the relative utility of Einhorn's nonlinear models is strongest in his job preference study, let us examine the

methodological differences between that study and the present one. One can isolate at least ten differences between the two studies:

(*a*) the kind of judges (non-professionals vs professionals);
(*b*) the type of task (job preferences vs differential diagnoses);
(*c*) the number of cues (2–6 vs 11);
(*d*) the values for each cue (discrete vs relatively continuous);
(*e*) the intercorrelations among the cues (contrived vs representative);
(*f*) the type of judgmental response (ranking vs rating);
(*g*) the resulting distribution of responses (rectangular vs normal);
(*h*) the number of cases (30 vs 861);
(*i*) the statistic used for comparing models (rank-order vs product-moment correlations);
(*j*) the use of "control" models (none vs two).

Clearly, one long-term goal of judgmental research must be the discovery of the effects of these variables upon the accuracy of various judgmental representations. In addition, future investigators who compare the Conjunctive or Disjunctive models with the Linear model should (*a*) use reasonably large and representative samples of experimental protocols, (*b*) include enough "control" models so as to make the findings as unambiguous as possible, and (*c*) pay special heed to the enormous problems involved in the use of models involving logarithmic or other nonlinear transformations of the original cues. Let us briefly consider each of these issues in turn.

Egon Brunswik (1947, 1955, 1956) has argued cogently for the use of "representative" research designs if experimental results are to be generalized outside the laboratory setting. On the other hand, it is often necessary to use a selected set of nonrepresentative protocols (e.g., "critical cases") in order to utilize some powerful statistical technique or to compare the predictions made by two or more highly correlated models. For example, in using an analysis-of-variance model for testing the significance of interaction effects (e.g., Hoffman, Slovic, & Rorer, 1968), it is typically necessary to use a contrived set of protocols in which all cues are orthogonal. And, in order to test for differences in the accuracy of highly correlated judgmental models (e.g., Linear vs Conjunctive), it may sometimes be necessary to utilize a set of protocols which, while statistically rare in nature, provide fine-grained illumination at just those points where the competing models generate diverse predictions. Fortunately, however, it is not necessary to purchase these special cases at the cost of nonrepresentative design. Instead, an ideal set of protocols should be both representative vis-a-vis the correlations among the cues (with the result that the judge is provided with as natural a judgmental problem as possible) *and* large enough to include a sizable subset of

"critical" protocols for which the cue values are orthogonal or otherwise constrained.

The case for adequate controls in experimental investigations has been made often enough, and well enough, that it should not need to be repeated once again. Yet, judgmental investigators seem particularly lax in constructing "control" models as bridges between the ones they particularly wish to compare. If a judge's responses are predicted better by the Conjunctive than the Linear model, is he really using a "multiple-cutting-score" strategy? As Hoffman (1960) originally pointed out when he coined the term "paramorphic" representation, we can never know with certainty the exact process used by a judge; all we can do is to compare alternative models, and hopefully discover one which predicts his responses as highly as their reliability permits. Consequently, before inferring that a judge is using a Conjunctive decision process, we must certainly rule out the possibility that he simply tends to minimize the differences among high (or low) cue values when making his judgments. Control strategies, such as the Logarithmic and the Exponential, could provide a gross test of this competing hypothesis.[5]

However, these two models, like the Conjunctive and Disjunctive approximations proposed by Einhorn, involve nonlinear transformations of the original cues and/or the judgmental responses. And, as Stevens (1968) has repeatedly emphasized, all such transformations lead to results which are completely arbitrary, unless the cues (and responses) are measured on a ratio scale. That is, without some natural zero point, the logarithms (and hence the accuracy of the Conjunctive model) will change with any linear change of scale. And, additionally, without some natural "upper bound" (the constant a_i in Table 1), the accuracy of the Disjunctive model will also be changed with any change of scale.[6] In

[5] By computer simulation, it was possible to construct hypothetical judges programmed to apply various strategies to the 861 MMPI profiles, and then to discover the extent to which each of the five models captured each of these known strategies. In one such simulation, a hypothetical judge programmed to use a Logarithmic strategy was better represented by the Conjunctive model ($R = .999$) than by the Linear model ($R = .840$). In a real study, an investigator might easily conclude that such a judge was using a multiple-cutting-score strategy if the Logarithmic model ($R = 1.00$) was not included as a control. As a second example, a hypothetical judge programmed to use an Exponential strategy was better represented by the Conjunctive model ($R = .988$) than by the Linear model ($R = .504$). Again, an unsuspecting investigator might conclude erroneously if the Exponential model ($R = 1.00$) was not included in his study.

[6] Additional analyses were carried out using different values of the constant a_i in the Disjunctive model. While the values of the multiple correlations varied with each new value of a_i, the conclusions derived from the analyses displayed in Tables 2–5 did not.

this sense, then, Einhorn's approximations of the Conjunctive and Disjunctive strategies, as well as the Logarithmic and Exponential models used in the present study, all lead to arbitrary findings, as long as the cues and responses are not measured on ratio scales. And, since few psychological variables are ever measured with such precision, Einhorn's models must be viewed as but crude first approximations to the sort of model needed to capture the notion of Conjunctive and Disjunctive decision strategies.

REFERENCES

ANDERSON, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817–818.

ANDERSON, N. H. Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, 1965, **70**, 394–400.

ANDERSON, N. H. Averaging model analysis of set-size effect in impression formation. *Journal of Experimental Psychology*, 1967, **75**, 158–165.

ANDERSON, N. H. Application of a linear-serial model to a personality-impression task using serial presentation. *Journal of Personality and Social Psychology*, 1968, **10**, 354–362.

ANDERSON, N. H. Comment on "An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment." *Psychological Bulletin*, 1969, **72**, 63–65.

BRUNSWIK, E. *Systematic and Representative Design of Psychological Experiments*. Berkeley: University of California Press, 1947.

BRUNSWIK, E. Representative design and probabilistic theory. *Psychological Review*, 1955, **62**, 236–242.

BRUNSWIK, E. *Perception and the Representative Design of Psychological Experiments. Berkeley*: University of California Press, 1956.

DAWES, R. M. Social selection based on multidimensional criteria. *Journal of Abnormal and Social Psychology*, 1964, **68**, 104–109.

EINHORN, H. J. The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 1970, **73**, 221–230.

EINHORN, H. J. Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance*, 1971, **6**, 1–27.

GOLDBERG, L. R. Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs*, 1965, **79**, (9, Whole No. 602).

GOLDBERG, L. R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 1968, **23**, 483–496.

GOLDBERG, L. R. The search for configural relationships in personality assessment: The diagnosis of psychosis vs. neurosis from the MMPI. *Multivariate Behavioral Research*, 1969, **4**, 523–536.

HAMMOND, K. R. Probabilistic functioning and the clinical method. *Psychological Review*, 1955, **62**, 255–262.

HAMMOND, K. R., HURSCH, C. J., & TODD, F. J. Analyzing the components of clinical inference. *Psychological Review*, 1964, **71**, 438–456.

Hammond, K. R., & Summers, D. A. Cognitive dependence on linear and nonlinear cues. *Psychological Review*, 1965, **72**, 215–224.

Hoffman, P. J. The paramorphic representation of clinical judgment. *Psychological Bulletin*, 1960, **57**, 116–131.

Hoffman, P. J. Cue-consistency and configurality in human judgment. In B. Kleinmuntz (Ed.), *Formal Representation of Human Judgment*. New York: Wiley & Sons, 1968. Pp. 53–90.

Hoffman, P. J., Slovic, P., & Rorer, L. G. An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin*, 1968, **69**, 338–349.

Meehl, P. E. A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *Journal of Counseling Psychology*, 1959, **6**, 102–109.

Naylor, J. C. Accuracy and variability of information sources as determiners of performance and source preference of decision makers. *Journal of Applied Psychology*, 1964, **48**, 43–49.

Naylor, J. C. Some comments on the accuracy and the validity of a cue variable. *Journal of Mathematical Psychology*, 1967, **4**, 154–161.

Naylor, J. C., & Schenck, E. A. The influence of cue redundancy upon the human inference process for tasks of varying degrees of predictability. *Organizational Behavior and Human Performance*, 1968, **3**, 47–61.

Naylor, J. C., & Wherry, R. J., Sr. The use of simulated stimuli and the "JAN" technique to capture and cluster the policies of raters. *Educational and Psychological Measurement*, 1965, **25**, 969–987.

Slovic, P. Cue-consistency and cue-utilization in judgment. *American Journal of Psychology*, 1966, **79**, 427–434.

Slovic, P. Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology*, 1969, **53**, 255–263.

Stevens, S. S. Measurement, statistics, and the schemapiric view. *Science*, 1968, **161**, 849–856.

Tucker, L. R. A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd. *Psychological Review*, 1964, **71**, 528–530.

Wiggins, N., & Hoffman, P. J. Three models of clinical judgment. *Journal of Abnormal Psychology*, 1968, **73**, 70–77.

Received: July 29, 1970

AMERICAN DOCUMENTATION INSTITUTE TABLES

TABLE A
A Comparison of the Multiple Correlations for Each of Five Models
Sample A (N = 92)

Judge	LIN	CON	DIS	LOG	EXP
1	.83*	.73	.56	.83	.72
2	.88	.80	.63	.88*	.78
3	.81	.75	.60	.82*	.73
4	.79*	.72	.57	.77	.72
5	.78	.71	.55	.79*	.70
6	.91	.85	.67	.91*	.82
7	.84*	.73	.63	.84	.73
8	.77	.78	.49	.81*	.73
9	.89*	.80	.69	.89	.79
10	.79	.79	.57	.80*	.75
11	.87	.87	.67	.90*	.82
12	.81	.72	.59	.82*	.71
13	.87*	.77	.67	.86	.76
14	.74*	.60	.60	.71	.63
15	.84*	.77	.66	.80	.82
16	.91*	.80	.67	.90	.80
17	.69*	.65	.51	.69	.66
18	.94	.90	.67	.95*	.86
19	.92	.86	.65	.92*	.82
20	.81	.80	.60	.84*	.76
21	.92*	.81	.72	.91	.81
22	.92	.86	.71	.93*	.83
23	.86	.78	.59	.87*	.75
24	.78	.75	.50	.80*	.71
25	.91	.84	.66	.92*	.82
26	.83	.78	.55	.84*	.75
27	.76*	.69	.64	.74	.69
28	.72	.77*	.51	.76	.72
29	.74	.77	.53	.78*	.73
Average	.83	.77	.61	.84*	.75
Composite	.92	.91	.73	.93*	.89
Criterion	.55	.57	.41	.57*	.55

* Best-fitting model.

TABLE B
A Comparison of the Multiple Correlations for Each of Five Models
Sample B (N = 77)

Judge	LIN	CON	DIS	LOG	EXP
1	.59	.55	.51	.61*	.54
2	.88*	.79	.70	.87	.79
3	.77*	.69	.61	.77	.66
4	.86*	.77	.78	.84	.79
5	.85*	.75	.68	.84	.75
6	.85*	.76	.75	.82	.79
7	.85	.81	.73	.86*	.79
8	.91	.84	.73	.92*	.82
9	.86	.80	.66	.87*	.76
10	.86	.81	.76	.86*	.80
11	.92	.86	.75	.93*	.84
12	.84*	.71	.74	.78	.76
13	.80*	.72	.70	.78	.72
14	.56	.57	.59*	.55	.52
15	.88*	.77	.77	.83	.83
16	.87*	.78	.75	.86	.79
17	.83*	.71	.79	.74	.81
18	.89	.82	.70	.89*	.79
19	.85	.80	.69	.86*	.77
20	.91	.85	.75	.91*	.83
21	.87	.79	.73	.88*	.78
22	.90	.86	.76	.91*	.85
23	.77	.77	.68	.79*	.71
24	.78*	.68	.66	.75	.71
25	.92	.85	.74	.92*	.83
26	.88	.80	.70	.89*	.78
27	.90*	.83	.76	.89	.82
28	.90*	.83	.71	.90	.81
29	.85*	.77	.75	.84	.79
Average	.84*	.77	.71	.83	.77
Composite	.93*	.91	.85	.93	.91
Criterion	.49*	.49	.40	.49	.49

* Best-fitting model.

TABLE C
A COMPARISON OF THE MULTIPLE CORRELATIONS FOR EACH OF FIVE MODELS
SAMPLE C ($N = 103$)

Judge	LIN	CON	DIS	LOG	EXP
1	.79*	.70	.61	.77	.70
2	.89*	.80	.74	.87	.82
3	.82*	.74	.63	.80	.75
4	.85*	.75	.70	.82	.77
5	.87*	.76	.67	.84	.78
6	.88*	.77	.72	.85	.81
7	.82*	.75	.70	.81	.75
8	.90	.84	.70	.91*	.81
9	.90	.84	.73	.91*	.81
10	.77	.74	.64	.77*	.73
11	.93	.88	.75	.93*	.86
12	.86*	.76	.64	.86	.75
13	.84*	.72	.65	.83	.72
14	.58*	.52	.55	.54	.52
15	.88*	.75	.68	.83	.80
16	.91*	.82	.76	.89	.83
17	.83*	.71	.65	.79	.75
18	.90*	.81	.67	.90	.79
19	.82	.78	.67	.83*	.77
20	.90*	.86	.77	.89	.85
21	.93*	.86	.77	.93	.86
22	.84	.79	.70	.84*	.77
23	.89*	.79	.66	.88	.80
24	.74	.72	.59	.74*	.70
25	.92	.88	.75	.92*	.86
26	.87	.81	.66	.88*	.79
27	.87*	.82	.73	.86	.82
28	.92*	.83	.69	.91	.83
29	.86	.79	.69	.86*	.78
Average	.85*	.78	.68	.84	.78
Composite	.94	.92	.80	.95*	.91
Criterion	.50*	.46	.48	.46	.50

* Best-fitting model.

TABLE D

A Comparison of the Multiple Correlations for Each of Five Models
Sample D ($N = 42$)

Judge	LIN	CON	DIS	LOG	EXP
1	.77	.83*	.75	.81	.74
2	.87	.88*	.76	.85	.85
3	.83*	.76	.81	.79	.80
4	.87*	.84	.76	.83	.84
5	.86*	.77	.74	.85	.78
6	.85	.86*	.83	.84	.84
7	.83	.85*	.72	.84	.78
8	.89*	.80	.75	.87	.81
9	.88*	.83	.74	.87	.82
10	.74	.68	.78*	.70	.74
11	.93	.86	.80	.93*	.85
12	.82*	.76	.67	.79	.79
13	.80*	.67	.63	.78	.71
14	.81	.84*	.76	.82	.77
15	.86	.89*	.78	.87	.84
16	.88	.90	.81	.91*	.82
17	.81*	.81	.69	.77	.77
18	.91*	.82	.70	.86	.85
19	.84*	.79	.76	.81	.81
20	.77*	.73	.70	.75	.74
21	.88*	.75	.72	.86	.76
22	.81*	.71	.70	.78	.73
23	.76*	.62	.71	.72	.66
24	.79*	.75	.63	.79	.76
25	.87	.81	.73	.87*	.80
26	.85*	.84	.73	.84	.82
27	.88*	.79	.72	.86	.81
28	.90*	.88	.80	.89	.86
29	.88*	.85	.79	.88	.85
Average	.84*	.80	.74	.83	.79
Composite	.93*	.91	.86	.92	.92
Criterion	.66*	.62	.57	.62	.66

* Best-fitting model.

TABLE E
A COMPARISON OF THE MULTIPLE CORRELATIONS FOR EACH OF FIVE MODELS
SAMPLE E (N = 181)

Judge	LIN	CON	DIS	LOG	EXP
1	.74*	.68	.66	.67	.73
2	.85*	.74	.71	.81	.77
3	.69*	.59	.58	.65	.62
4	.68*	.54	.61	.58	.64
5	.82*	.72	.67	.78	.76
6	.78*	.65	.76	.66	.77
7	.70*	.63	.60	.68	.65
8	.84*	.79	.67	.84	.77
9	.82*	.74	.68	.81	.74
10	.76*	.70	.65	.74	.71
11	.86*	.79	.69	.84	.79
12	.78*	.67	.61	.75	.69
13	.76*	.65	.61	.73	.67
14	.59	.47	.60*	.54	.51
15	.84*	.76	.73	.78	.81
16	.85*	.76	.72	.82	.79
17	.71*	.54	.58	.62	.64
18	.88*	.71	.74	.75	.83
19	.84*	.76	.70	.81	.77
20	.80*	.71	.69	.77	.73
21	.88*	.79	.74	.87	.79
22	.82*	.74	.68	.80	.74
23	.77*	.67	.66	.72	.70
24	.66*	.59	.65	.61	.65
25	.83*	.77	.65	.82	.76
26	.72*	.59	.54	.68	.61
27	.82*	.73	.67	.81	.74
28	.87*	.69	.73	.76	.81
29	.80*	.74	.71	.78	.75
Average	.79*	.69	.67	.74	.72
Composite	.88*	.83	.80	.85	.86
Criterion	.55*	.49	.53	.49	.55

* Best-fitting model.

TABLE F
A Comparison of the Multiple Correlations for Each of Five Models
Sample F (N = 166)

Judge	LIN	CON	DIS	LOG	EXP
1	.72	.73*	.66	.71	.71
2	.84*	.75	.71	.82	.77
3	.79*	.72	.65	.78	.72
4	.81*	.72	.75	.75	.77
5	.80*	.73	.65	.79	.73
6	.81*	.72	.71	.78	.74
7	.71*	.58	.66	.63	.66
8	.87*	.80	.68	.87	.79
9	.85*	.79	.70	.83	.79
10	.88*	.88	.75	.88	.85
11	.92*	.86	.77	.90	.85
12	.77*	.65	.64	.71	.71
13	.77*	.68	.67	.75	.70
14	.71*	.59	.67	.67	.63
15	.88*	.80	.78	.81	.85
16	.86*	.78	.73	.84	.79
17	.83*	.69	.78	.72	.81
18	.89*	.78	.74	.84	.83
19	.78*	.72	.67	.77	.72
20	.86*	.77	.72	.83	.78
21	.90*	.84	.76	.88	.84
22	.80*	.76	.71	.79	.77
23	.80*	.68	.69	.76	.72
24	.81*	.67	.65	.76	.72
25	.92*	.88	.76	.92	.86
26	.83*	.77	.69	.81	.77
27	.69*	.59	.65	.63	.63
28	.91*	.84	.76	.90	.85
29	.85*	.78	.74	.83	.79
Average	.82*	.74	.71	.79	.76
Composite	.91*	.88	.81	.90	.89
Criterion	.49*	.45	.46	.45	.49

* Best-fitting model.

TABLE G
A COMPARISON OF THE MULTIPLE CORRELATIONS FOR EACH OF FIVE MODELS
SAMPLE G (N = 200)

Judge	LIN	CON	DIS	LOG	EXP
1	.81*	.74	.73	.79	.76
2	.87*	.77	.72	.85	.78
3	.79*	.72	.68	.78	.73
4	.81*	.72	.68	.78	.73
5	.84*	.75	.68	.82	.75
6	.89*	.83	.77	.88	.83
7	.85*	.76	.72	.83	.76
8	.87*	.81	.72	.87	.80
9	.90*	.83	.75	.89	.82
10	.84	.82	.71	.85*	.81
11	.92*	.87	.76	.91	.85
12	.82*	.75	.67	.81	.75
13	.78*	.70	.65	.77	.69
14	.66*	.50	.58	.61	.55
15	.87*	.78	.73	.84	.80
16	.89*	.82	.75	.88	.82
17	.83*	.70	.69	.78	.74
18	.89*	.80	.68	.88	.80
19	.83*	.78	.69	.83	.77
20	.84*	.76	.70	.83	.76
21	.92*	.86	.77	.91	.84
22	.88*	.81	.74	.87	.81
23	.77*	.63	.67	.73	.68
24	.77*	.72	.64	.77	.71
25	.89*	.83	.74	.89	.82
26	.83*	.73	.62	.81	.73
27	.89*	.85	.74	.89	.84
28	.88*	.79	.71	.87	.79
29	.84*	.77	.70	.83	.78
Average	.84*	.77	.70	.83	.77
Composite	.93*	.90	.83	.92	.91
Criterion	.65*	.62	.61	.62	.65

* Best-fitting model.