NEW DIRECTIONS FOR METHODOLOGY OF SOCIAL AND BEHAVIORAL SCIENCE

A Quarterly Sourcebook
Donald W. Fiske, Editor-In-Chief

Number 4, 1980

Fallible Judgment in Behavioral Research

Richard A. Shweder
Guest Editor

Jossey-Bass Inc., Publishers
San Francisco • Washington • London

Likeness in meaning does not predict covariation between behaviors; thus, it is hazardous to substitute propositions about language for propositions about the world.

The Systematic Distortion Hypothesis

Richard A. Shweder
Roy G. D'Andrade

Personality researchers who rely on memory-based assessment procedures (inventories, checklists, questionnaires) to study the organization of individual differences tend to discover generalized personality traits, coherent syndromes, and simple main effects (Block, 1965; Cattell, 1946; LaForge and Suczek, 1955; Norman, 1963; Smith, 1967). The systematic distortion hypothesis (D'Andrade, 1965, 1973, 1974; Shweder, 1975, 1977a, b; Shweder and D'Andrade, 1979b) suggests that much of this memory-based evidence in support of global personality trait structure is artifactual. In this chapter we examine the systematic distortion hypothesis and discuss its implications for our understanding of personality structure, human judgment, and implicit personality theory.

The Systematic Distortion Hypothesis

The systematic distortion hypothesis states that under difficult memory conditions judges on personality inventories, rating forms, and questionnaire

Research on the Brief Psychiatric Rating Scale and videotape of family interaction was carried out with support from the National Institute of Mental Health under Grant No. MN19864 to D'Andrade. An earlier version of this chapter was entitled "Personality or Culture?" Preparation of this chapter was made possible in part by a grant from the Public Health Service, PHS 5S07RR07029-13, to Shweder.

interviews infer what "must" have happened from their general beliefs about what the world is like and/or find it easier to retrieve conceptually related memory items. The hypothesis further states that our general beliefs about what the world is like in the area of personality (our implicit personality theories) tend to be inaccurate with respect to how behaviors covary, confusing "what is like what" with "what goes with what"; therefore interbehavior correlations derived from memory-based assessment procedures cannot be considered valid evidence for the objective existence of proposed personality traits, factors, or syndromes.

Informally stated, the systematic distortion hypothesis suggests that judges on memory-based personality procedures are prone to a cognitive illusion in which "propositions about language" are confused with "propositions about the world" (D'Andrade, 1965, p. 215) and likeness in meaning mistaken for co-occurrence likelihood (Shweder, 1977b; also see Chapman, 1967; Chapman and Chapman, 1967, 1969).

One way to test the systematic distortion hypothesis is to compare the degrees of correspondence between interbehavior patterns of association (for example, measures of correlation, distance, and mutual substitutability) derived from similarity of meaning judgments, memory-based hypothesis predicts that correlations *between behaviors* in memory-based personality ratings are not accurate reports about the interbehavior correlations found in actual behavior but instead reflect the degree to which the labels for the behaviors are similar in meaning.

Reproducing Memory-Based Rating Structures from Similarity of Meaning Judgments

During the last fifteen years a number of researchers have discovered that the taxonomic categories, factors, and dimensions (such as "character strength," "permissiveness," "emotional stability," "ego-resilience") induced from correlational patterns of response equivalence on standard memory-based personality instruments can also be derived by asking subjects how the items on the test are "similar in meaning" (D'Andrade, 1965, 1974, Ebbesen and Allen, 1977; Mulaik, 1964; Shweder, 1975, 1977a; see Shweder and D'Andrade, 1979b, for a review). To illustrate this procedure we have reproduced the correlational structure of a set of sixteen rating scales used in psychiatric diagnosis. The scales were developed by J. E. Overall and his associates, and are used primarily by psychiatrists and other mental health professionals for rating hospitalized patients (Overall and Hollister, 1968). Ratings are made at the end of a standard psychiatric intake interview. Reliabilities for the scales range from .56 to .86 (Overall and Gorham, 1962).

Based on cluster and factor analyses of the intercorrelations of the sixteen scales, Overall finds four "syndromes" or clusters of scales: "depressive disturbance," "thinking disturbance," "paranoid interpersonal disturbance," and "withdrawal retardation." The scales for these syndromes and factor analytic loadings, found in a number of samples, are given in Table 1.

Table 1. Brief Psychiatric Rating Scale Factor Loadings from Five Samples

	Depressive	*Thinking*	*Interpersonal*	*Withdrawal*
	Disturbance	*Disturbance*	*Disturbance*	*Retardation*

Table 1. Brief Psychiatric Rating Scale Factor Loadings from Five Samples

		Depressive Disturbance				*Thinking Disturbance*				*Interpersonal Disturbance*				*Withdrawal Retardation*			
		VA	SN	JN	HP	VA	SN	JN	HP	VA	SN	JN	HP	VA	SN	JN	HP
Guilt	5	41	46	81	73	19	-03	00	09	02	16	-05	04	-11	09	-03	-09
Depression	9	83	50	54	80	-08	-08	-19	06	-09	08	-07	-02	-11	37	33	05
Somatic	1	70	77	56	25	04	04	-02	10	09	10	-05	04	05	-01	-03	-05
Anxiety	2	68	79	83	71	11	-02	03	24	26	26	06	22	-18	-04	-11	-11
Tension	6	34	21	42	46	13	10	17	12	24	22	15	33	04	19	08	03
Disorganized ideas	4	-25	00	07	-06	32	33	67	46	13	34	-07	22	49	35	42	21
Odd ideas	15	-19	-02	08	-04	73	58	84	82	38	64	-06	18	19	15	06	-08
Hallucinations	12	-01	02	03	-02	80	79	72	82	18	20	-24	-16	22	25	27	01
Grandiosity	8	-35	-17	-01	-29	27	08	45	27	24	44	29	32	02	-13	-13	-11
Hostility	10	03	00	15	-02	-21	-32	33	-03	85	84	79	88	-04	04	02	-06
Suspicious	11	-03	-02	23	04	19	-04	53	30	84	87	33	79	08	00	-04	-13
Uncooperative	14	-02	-16	-03	09	-17	-09	20	-13	36	28	67	45	49	53	50	57
Withdrawn	3	04	-10	-03	02	-06	-01	06	02	-02	-01	08	00	88	85	87	90
Unresponsive	16	-07	-05	-02	00	-06	-01	-02	05	-21	00	-05	-05	76	87	88	82
Slow	13	42	19	04	05	-07	-06	-22	-07	-22	-14	-16	-29	41	74	74	65
Odd gesture	7	03	01	06	-01	15	08	17	02	20	11	00	01	62	54	31	53

Source: Overall, Hollister, and Pichot, 1967.

VA = Veterans Hospital Drug Screening Data N = 725
SN = Senior Nurse Ratings N = 549
JN = Junior Nurse Ratings N = 549
HP = Hospital Psychiatrists Ratings N = 549

In order to reproduce these syndromes or clusters of scales with similarity ratings made by laymen, the semi-technical psychiatric phrasing of the original scales was translated into ordinary language. The translated scales are presented in Figure 1. Ten university undergraduates were asked to rate all pairs of scales with regard to their similarity in meaning on a +100 to −100 scale, where +100 was defined as "completely identical in meaning" and −100 was defined as "completely opposite in meaning," and zero was defined as "unrelated in meaning."

The matrices for the intercorrelations of the psychiatrists' patient ratings and the undergraduates' similarity judgments are correlated .66. U-Statistic cluster analyses (D'Andrade, 1978) of the rating scales and the similarity judgments yield very similar results (see Tables 2 and 3). Both show the four "syndromes" of anxious depression, thinking disorder, paranoid interpersonal disturbance and withdrawal retardation. Two scales change clusters in the analysis of similarity ratings. The scale labelled "uncooperative" moves from the withdrawal retardation cluster in the psychiatrists' patient ratings to the paranoid interpersonal disturbance cluster in the undergraduate similarity ratings, while the scale labelled "odd gestures" moves from the withdrawal retardation cluster to the thinking disturbance cluster. Similarity of meaning judgments yield patterns of association that are similar to the patterns found in the intercorrelations of ratings.

Accurate Reflection or Systematic Distortion?

The correspondence between similarity of meaning structures and memory-based personality rating structures is compatible with two quite

Figure 1. Short Phrases from Ordinary Language Translation of the Brief Psychiatric Rating Scale

1. concern about bodily health; health as a major problem
2. worry and anxiety about things which have happened or might happen
3. withdrawn and not emotionally involved
4. confused and disorganized ideas and ways of talking
5. guilty or remorseful feelings; concern about things that might have been done wrong
6. physically tense and jittery
7. odd gestures, facial expressions, ways of moving
8. exaggerated idea of self-importance and belief in own unusual ability
9. sad, depressed, despondent
10. full of hostile, disdainful, and belligerent feelings toward other people
11. thinks other people are (or might have been) against him or out to hurt him
12. has visions; sees or hears things that other people do not see or hear
13. slowed down in thinking, talking, or moving
14. uncooperative, unfriendly, and resentful
15. odd or strange ways of thinking
16. unresponsive and almost without emotional reactions

Table 2. Rated Behavior Structure. Correlations in Psychiatrists' Assessments of 549 Patients Using Overall's Brief Psychiatric Rating Scale

	0	1	2	6	5	9	12	8	4	15	11	10	14	7	3	16	13
Somatic concerns	1	100	21	9	22	15	1	1	3	14	7	4	−11	2	−7	3	−9
Anxiety	2	21	100	49	32	41	15	−8	5	18	30	15	−8	−3	−6	−9	−15
Physically tense	6	9	49	100	26	25	3	5	14	13	22	24	17	15	10	−6	−16
Guilt feelings	5	22	32	26	100	40	2	−4	7	10	5	5	−11	−3	−8	−3	−9
Depressed	9	15	41	25	40	100	5	−22	−5	1	4	−2	−6	−4	5	3	16
Hallucinations	12	1	15	3	2	5	100	8	17	35	13	−15	−14	2	2	3	5
Grandiosity	8	1	−8	5	−4	−22	8	100	27	39	19	28	6	1	−8	−9	−27
Disorganized ideas	4	3	5	14	7	−5	17	27	100	50	23	17	12	20	23	13	1
Unusual thoughts	15	14	18	13	10	1	35	39	50	100	33	14	6	−3	−6	−1	−22
Suspicious	11	7	30	22	5	4	13	19	23	33	100	50	14	−13	−10	−11	−25
Hostile	10	4	15	24	5	−2	−15	28	17	14	50	100	31	−1	−7	−7	−29
Uncooperative	14	−11	−8	17	−11	6	−14	6	12	6	14	31	100	34	48	26	27
Odd gestures	7	2	−3	15	−3	−4	2	1	20	−3	−13	1	34	100	40	33	29
Withdrawn	3	−7	−6	10	8	5	2	−8	23	−6	−10	−7	48	40	100	62	51
Unresponsive	16	3	−9	−6	−3	3	3	−9	13	−1	−11	−7	26	33	62	100	35
Physically slow	13	−9	−15	−16	−9	16	5	−27	1	−22	−25	−29	27	29	51	36	100

Table 3. Conceptual Similarity Structure. Mean Similarity Ratings by 10 Informants (100 'always go together', 0 'unrelated', – 100 'never go together') for Modified Version of Overall's Brief Psychiatric Rating Scale

		1	2	6	5	9	12	8	4	15	7	11	10	14	3	16	13
Somatic concerns	1	100	60	20	– 5	30	– 35	– 5	– 50	– 10	– 30	20	– 10	– 65	– 25	– 55	– 90
Anxiety	2	60	100	80	70	40	10	– 55	– 10	35	60	60	10	0	– 15	– 40	– 55
Physically tense	6	20	80	100	15	– 65	10	– 10	40	15	– 25	60	15	10	– 70	– 75	– 95
Guilt feelings	5	– 5	70	15	100	80	– 15	– 50	5	0	0	5	– 35	– 30	– 20	– 75	0
Depressed	9	30	40	– 65	80	100	– 15	– 70	10	0	0	– 10	– 50	– 20	55	20	35
Hallucinations	12	– 35	10	10	– 15	– 15	100	50	40	70	30	35	5	– 25	5	– 55	– 50
Grandiosity	8	– 5	– 55	– 10	– 50	– 70	50	100	– 10	60	5	40	55	10	– 60	– 65	– 90
Disorganized ideas	4	– 50	– 10	40	5	10	40	– 10	100	80	65	25	10	– 5	20	10	5
Unusual thoughts	15	– 10	35	15	0	0	70	60	80	100	75	45	20	25	0	– 15	10
Odd gestures	7	– 30	60	– 25	0	0	30	5	65	75	100	10	0	– 35	0	– 5	0
Suspicious	11	20	60	60	5	– 10	35	40	25	45	10	100	70	60	– 5	– 65	– 60
Hostile	10	– 11	10	15	– 35	– 50	5	55	10	20	0	70	100	70	– 45	– 70	– 45
Uncooperative	14	– 65	0	10	– 30	– 20	– 25	10	– 5	25	– 35	60	70	100	40	0	– 15
Withdrawn	3	– 25	– 15	– 70	– 20	55	5	– 60	20	0	0	– 5	– 45	40	100	100	35
Unresponsive	16	– 55	– 40	– 75	– 75	20	– 55	– 65	10	– 15	– 5	– 65	– 70	0	100	100	70
Physically slow	13	– 90	– 55	– 95	0	35	– 50	– 90	5	10	0	– 60	– 45	– 15	35	70	100

divergent hypotheses. On the one hand, it is conceivable that semantic structures (patterns of conceptual association) and memory-based rating structures correspond because both are accurate reflections or encodings of what actually correlates with what in the real world. On the other hand, it is conceivable that memory-based personality ratings are systematically distorted in the direction of preexisting ideas about "what is like what" and that neither the semantic structures nor the rating structures tell us much about which behaviors correlate with each other in actual conduct.

One way to choose between the "accurate reflection" and "systematic distortion" hypotheses is to assess the degree of correspondence between the interbehavior patterns of association (clusters, traits, syndromes, or factors) derived from personality ratings and/or similarity of meaning judgments with the interbehavior patterns of association derived from a reasonably objective performance standard, such as reliable, on-line scorings of conduct. In general, comparisons of this type have supported the "systematic distortion" hypothesis (D'Andrade, 1973, 1974; Newcomb, 1929, 1931; Shweder, 1975, 1977a, b). What correlates with what in memory-based personality ratings corresponds to "what is like what" in similarity of meaning judgments and neither the rating structure nor the semantic structure tells us much about what actually correlates with what in the real world. As an illustration consider the following analysis of a thirty minute videotape of natural unstaged interaction among four members of a family.

Systematic Distortion: An Illustration

The material to be analyzed was taken from a nationally broadcast documentary series presented over public television. For over a year the ordinary and extraordinary events in the life of a white upper middle class California family were videotaped. From this record, twelve hour-length shows were produced. For this study, thirty minutes were selected from the last hour in which Pat and Bill, the mother and father of the family, and Lance and Delilah, the oldest son and the oldest daughter, appeared.

For the purpose of comparing what correlated with what in actual behavior and in memory-based ratings, it was important to select a set of relatively unambiguous everyday terms for describing interpersonal behavior. It was also important to select terms whose application occurred frequently enough to establish differences between the four people or "actors" to be scored and rated. A preliminary list of forty-five terms was taken from Osgood's semantic analysis of interpersonal behavior terms (Osgood, 1970). Using the transcripts and the videotape, two scorers were asked to determine each time an actor completed an act, and then to check on Osgood's list the term or terms they thought described the behavior. Using this procedure, the sixteen highest frequency terms were selected.

The next step was to have a set of on-line, immediate scorings made by three coders using just the sixteen terms. Again the scorers were asked to determine the boundaries of each act using the transcript and viewing the

tape, and then check on the list of sixteen terms the term or terms which characterized each act. On the average, the three scorers coded three hundred and sixty-nine acts following this procedure. Informal inspection of the coded transcripts indicates that the raters agreed in most cases on the segmentation of particular acts, apparently because the turn-taking system of natural conversation functions as an act-segmenting device.

Reliabilities were derived from the average of the product-moment correlations between pairs of scorers for each category of interpersonal behavior using percentages of acts across actors. The correlations were averaged for the three pairs of scorers using Fisher's r to Z transformation. The mean correlations are presented in Table 4.

Examination of the reliability figure scores across the four actors indicated that two of the categories were unreliable: "explain," with a mean r of −.42 and "support," with a mean r of −.40. For three other categories, the percents of acts were identical for all actors: "lead," "warn," and "be courteous." These five categories were dropped from the analysis.

For the memory-based rating condition, twenty university undergraduates viewed the tape. No more than three subjects viewed the tape at the same time. When two or more subjects viewed the tape together they were instructed not to talk about it and to refrain from showing any reaction. As each of the four actors appeared on the screen they were identified by the experimenter. Subjects were told that they would be given a questionnaire about the behavior and actions of the actors after the tape was finished. The questionnaire was administered immediately upon completion of the tape. On each page of the questionnaire the actor's name was written along with scales

Table 4. Product Moment Correlations Between Pairs of Scorers Across Actors by Behavior Category

Category	*Scorer 1 and 2*	*Scorer 1 and 3*	*Scorer 2 and 3*	*Mean r*
1. inform	.54	.23	.73	.53
2. question	.99	.98	.99	.99
3. explain	.12	.24	−.93	−.40
4. joke	.84	.73	.97	.89
5. criticize	.97	.86	.81	.91
6. agree	.83	.24	.04	.45
7. disagree	.70	.99	.79	.91
8. advise	.74	.53	.95	.82
9. ridicule	.99	.97	.98	.98
10. suggest	−.22	.79	.21	.34
11. praise	.53	.74	.94	.80
12. lead	.69	.79	.99	.91
13. warn	.81	.96	.93	.92
14. support	.26	−.23	−.03	−.00
15. comply	.61	.71	.77	.70
16. be courteous	.46	.59	.90	.71

for each behavior category. Raters were asked "How much does [so-and-so] do the following [for example, inform others]?" Ratings were made on a seven point scale (1 = not at all; 7 = a lot).

For each judge in the memory-based rating condition, Kendall's tau was computed for all pairs of categories across actors. To compute tau for a pair of categories, each judge's ratings were used to obtain a rank ordering of the four actors on both categories, and tau was then computed from the two rank orderings. Tau coefficients were then averaged across the twenty judges for each pair of categories. The matrix of mean tau coefficients is presented in Table 5.

To obtain a correlation matrix for the immediate on-line scorings, mean percentage figures were used to rank order actors on each category. Tau coefficients were computed from the rank orders for all pairs of categories. The results are also presented in Table 5.

By inspection it can be seen that the pattern of association for the memory-based ratings is quite different than the pattern of association for the immediate scorings. The lines in Table 5 represent the major clusters discovered in a U-Statistic cluster analysis of similarity of meaning judgments for these categories. The product moment correlation between the matrix for memory-based ratings and the matrix for immediate scorings of behavior (Table 5) is only .22.

In order to compare the interbehavior correlations derived from both immediate, on-line scoring, and memory-based ratings with judgments of semantic similarity, judgments were obtained from ten undergraduates for all pairs of eleven categories on a scale running from "identical in meaning" to "completely opposite in meaning." The intervals of the scale ran from + 100 to - 100. The mean judgments are presented in Table 6.

The matrix for the similarity of meaning judgments is similar in its pattern of associations to the matrix for the memory-based ratings (r = .75), but

Figure 2. Degrees of Correspondence (Pearson r) Between Correlational Structures Derived from Similarity of Meaning Judgments (Semantic Structure), Memory-Based Ratings (Rating Structure), and Immediate Scorings (Behavioral Structure) in Videotape Study

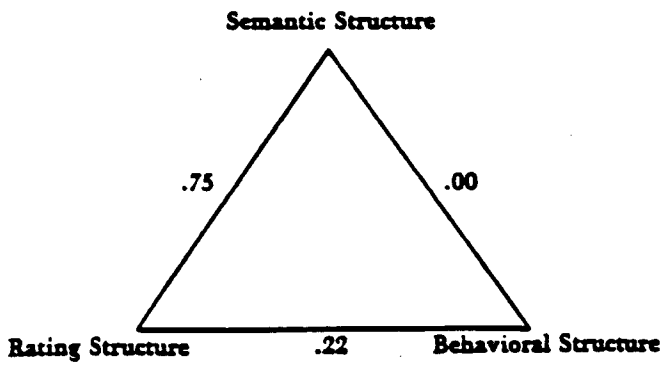

Table 5. Mean Tau Coefficients for Actor Rank Orders Derived from Memory-Based Ratings (Upper Half Matrix) and Tau Coefficients for Actor Rank Orders Derived from Immediate Scorings (Lower Half Matrix)

	ag	*co*	*pr*	*ad*	*in*	*su*	*qu*	*cr*	*di*	*jo*	*ri*
Agree	•	.28	.29	.07	.11	.06	−.01	−.31	−.42	.16	−.24
Comply	−.67	•	.39	.07	.11	.12	−.11	−.31	−.18	−.25	−.50
Praise	.00	.33	•	.08	−.05	.03	−.06	−.46	−.31	.21	−.27
Advise	.33	−.67	−.67	•	.42	.51	.24	.10	.05	−.33	−.03
Inform	.00	−.33	.33	.00	•	.37	.14	.00	−.10	−.21	−.24
Suggest	−.67	.33	−.33	.00	−.33	•	.11	.13	−.02	−.49	−.29
Question	−.67	.33	.33	.00	.33	.33	•	.12	−.01	−.19	−.13
Criticize	.00	−.33	−1.00	.67	−.33	.33	−.33	•	.59	−.30	.39
Disagree	1.00	−.67	.00	.33	.00	−.67	−.67	.00	•	−.11	.45
Joke	.00	.33	.33	−.67	−.33	−.33	−.33	−.33	.00	•	.17
Ridicule	.33	.00	−.67	.33	−.67	.00	−.67	.67	.33	.00	•

Agree
Comply
Praise

Advise
Inform
Suggest

Question
Criticize
Disagree

Joke
Ridicule

not a
(And,
discov
actual
the "s
bias i
reflec
the re

Impli

global
relatic
distor
hypot
instea

assess
under
are, o
ences
effect

exam
duces
evide
and T
cover

Table 6. Mean Conceptual Similarity Ratings, 10 Subjects

	ag	*co*	*pr*	*ad*	*in*	*su*	*qu*	*cr*	*di*	*jo*	*ri*
Agree	•	68	60	-16	-16	04	-36	-60	-32	-44	-56
Comply	68	•	36	04	-12	-32	-36	-60	-56	-56	-44
Praise	60	36	•	-28	-36	12	-32	-40	-52	00	-44
Advise	-16	04	-28	•	76	88	-04	-08	-36	-44	-36
Inform	-16	-12	-36	76	•	64	-12	-12	-28	-52	-36
Suggest	04	-32	12	88	64	•	-16	-28	16	-36	-44
Question	-36	-36	-32	-04	-12	-16	•	44	72	-40	-20
Criticize	-60	-60	-40	-08	-12	-28	44	•	44	-04	80
Disagree	-32	-56	-52	-36	-28	16	72	44	•	-16	24
Joke	-44	-56	00	-44	-52	-36	-40	-04	-16	•	88
Ridicule	-56	-44	-44	-36	-36	-44	-20	80	24	88	•

not at all like the matrix for the immediate scorings of behavior ($r = .00$). (And, as already noted, the latter two matrices correlate only .22.) What we discover in these various comparisons of memory-based rating structure, actual behavior structure, and conceptual similarity structure is predicted by the "systematic distortion" hypothesis. That is, there appears to be a systematic bias in memory. Correlations between categories in memory-based ratings reflect the degree to which the categories are similar in meaning, rather than the relationships found in immediately scored behavior. See Figure 2.

Implications for Studies of Personality Structure

In the personality literature, much of the evidence for the existence of global personality traits, syndromes, or factors is based on interbehavior correlational evidence derived from memory-based ratings. The systematic distortion hypothesis suggests that such evidence cannot be trusted. The hypothesis challenges us to construct a portrait of personality structure based instead on on-line records of conduct.

It seems likely that a shift in research methods from memory-based assessment procedures to behavioral observational techniques will alter our understanding of personality structure. Most researchers who utilize immediate, on-line scorings techniques to study the organization of individual differences tend to discover complex person-context-response mode interaction effects and undramatic interbehavior correlations.

Broad, empirically homogeneous multi-item traits or syndromes (for example, extrovert: likes parties, feels at ease talking before a group, introduces himself to strangers) are difficult to induce from behavior observational evidence (Mischel, 1968; Moos, 1969; Newcomb, 1929; Raush, Dittmann, and Taylor, 1959; Sears, 1963; Yarrow and Waxler, 1976). Instead, one discovers that *minor changes in context, task, or response mode produce major changes in*

individual difference rankings (Campbell and Fiske, 1959; Cronbach, 1975; Fiske, 1978; Mischel, 1968; Moos, 1969; Newcomb, 1929; Raush, Dittmann, and Taylor, 1959; Shweder, 1979a; Slovic, 1972b).

The Whitings' findings (1975, p. 163, for example,) are representative of available behavior observational evidence on the generalization of individual differences across *contexts*. In a study of children's social behavior to peers, infants, and adults, the highest level of cross-contextual generalization for a purported trait of behavior was .29 (Pearson r; n = 134). For the response tendency "nurturance" the degree of generalization was .05. Knowing that a child is relatively more nurturant than others *to adults*, for example, tells you little about whether he will be relatively more nuturant than others *to peers*.

Sears' findings (1963) are representative of available behavior observational evidence on the generalization of individual differences across *response modes*. In a study of "dependency" in children, Sears discovered that, on the average, "similar" response modes (such as seeks attention, seeks help, seeks physical nearness) intercorrelated .21. Children who seek attention more than others from their mothers are not much more likely than other children to "cling to their mothers' apron strings."

How sensitive are individual difference rankings to variations in context (time, place, personnel), task, and response mode? At the moment this is unclear (although for a despairing view see Cronbach, 1975). What does seem clear is that the world of individual differences in behavior is not organized in the way envisioned by personality trait theorists. Individual differences exist, of course, but they do not seem to generalize widely across similar contexts, tasks, or response modes. As Slovic (1972b) remarks in his review of the evidence against the existence of the purported trait, "risk-taker," as a generalized response tendency of individuals: "Only those tasks highly similar in structure and involving the same sorts of payoffs (all financial, all social) have shown any generality . . . and, as similarity decreases, these intertask correlations rapidly decrease . . ." (p. 128).

In summary, when one has a reasonably objective performance standard (such as on-line scorings of a videotape) on individual differences across comparable contexts, tasks, and response modes, what one discovers is a world of complex statistical interaction effects, multiple necessary conditions, and insubstantial intercorrelations among events. We may well live in a world where a relevant disposition is the tendency for middle-aged men to get angry when extravagantly dressed middle-aged women cut in front of them in line, but this regularity in someone's conduct may tell us little about whether he is more likely than others to get angry when contradicted in an argument at a scientific meeting.

Notably, this complex organization of individual differences is *not* encoded in implicit personality theory, and is rarely reported on personality rating forms, inventories, or questionnaires. Quite the contrary, our everyday theories of what goes with what in personality ("friendly" and "smiles a lot" go together; "gentle" and "managerial" exclude one another) portray a world of neat clusters, simple main effects, and widely generalized regularities. It is that conceptually coherent, neatly integrated portrait of the organization of

individual differences that judges seem to report on memory-based personality tests (see D'Andrade, 1965; LaForge and Suczek, 1955; Mulaik, 1964; Norman, 1963; Passini and Norman, 1966; Shweder, 1975).

Implications for the Study of Human Judgment

The systematic distortion hypothesis states that judges on personality instruments have difficulty remembering the correlational structure of observed behavioral events. One implication of the hypothesis, supported by some recent research on "probability learning," is that human learning mechanisms do not guarantee insight into the "statistical structure of sequences of events" (Estes, 1976a, p. 51). (See also Einhorn and Hogarth, 1978.) Individual observers are not skillful at arriving at a vertical understanding of the contingencies in their environment.

This is not only true for event sequences with complex statistical structures; it seems that observers readily comprehend the contingent structure of experience only under quite special circumstances (Estes, 1976a, b). As one study of correlational thinking concludes, "those who receive information on a trial by trial basis, as it usually occurs in the real world, generally fail to assess adequately the degree of relationship present" (Ward and Jenkins, 1965, p. 240; also see Einhorn and Hogarth, 1978; Jenkins and Ward, 1965; Shweder, 1977c; Slovic, 1972a; Smedslund, 1963).

Recent research on probability learning contains three related messages: (1) Judges often assess "the probability of an event by the ease with which instances or occurrences can be brought to mind" (Tversky and Kahneman, 1974, p. 1127; see also 1973); (2) Event categories are sometimes encoded, accessed, and retrieved, that is, "brought to mind," according to principles unrelated to event probabilities, thereby resulting in erroneous estimations of event likelihoods; (3) Two of the potentially hazardous principles for encoding, accessing, and retrieving event categories are "relative frequency" (Estes, 1976a, b) and "verbal association strength" or "conceptual affiliation" (Chapman, 1967; Chapman and Chapman, 1967, 1969).

Tversky and Kahneman (1974), for example, in a discussion of human predictive behavior, describe a number of techniques relied on by judges to estimate event likelihoods. One very general technique is known as the "availability heuristic": people assess "the probability of an event by the ease with which instances or occurrences can be brought to mind," as noted earlier.

Tversky and Kahneman argue that the availability heuristic is relied on in estimation tasks because "lifelong experience has taught us that, in general, instances of large classes are recalled better and faster than instances of less frequent classes; that likely occurrences are easier to imagine than unlikely ones . . ." (1974, p. 1128). Tversky and Kahneman are quick to point out, however, that there are many influences on the ease with which instances of an event class can be brought to mind, and since many of these influences are unrelated to the probability of an event, the availability heuristic has the potential to produce massive and systematic biases in judgment.

Just how hazardous it is to rely on the availability heuristic is made

apparent by Estes (1976a) who suggests that more likely occurrences are not easier to bring to mind than less likely occurrences unless the more likely occurrence also happens to be the more frequent occurrence. As Estes notes: human predictive behavior is (p. 45) "not a probability estimate but rather a record in memory of past frequencies of events": *subjects invariably rely on relative frequency information to estimate the probabilities of events even when relative frequency information is unrelated to event likelihoods*. Certain classes of events that are relevant for probability estimation, that is, the number of times an event that might have occurred did *not* occur, are just not processed. It is not more *likely* events that are easier to retrieve but more *frequent* ones. The availability heuristic would seem to be a "rational" technique for estimating event likelihoods only if the relative frequency of an event happens to correlate with its likelihood.

It can be dangerous to rely on an accurate relative frequency sensor when estimating the likelihood of an event. For example, in one experiment Estes (1976a) permitted subjects to observe a contest in which A was pitted against B 100 times with a .75 probability of success (A won 75 times) while C was pitted against D 200 times with a .50 probability of success (C won 100 times). Subjects were then asked to predict the results of an uncertain event: who will win if A is pitted against C? As Estes (1976a) notes: "it seems clear that if he bases his prediction on rational grounds he should predict A over C, and at worst, if he is unable to transfer acquired information to the test situation, he should mentally toss a coin and predict A or C with equal probabilities" (p. 43).

Neither "rational" prediction occurred. Despite the fact that A had won a greater proportion of contests against its opponents (75 out of 100) than C had won against its opponents (100 out of 200) in every instance subjects predicted the winner of the A versus C contest to be C, "the stimulus that accumulated the largest number of wins during the observation series regardless of its (past) probability of winning or losing" (Estes, 1976a, p. 44). Subjects failed to process information about two rather large classes of events, the number of *losses* incurred by A and C, that is, the number of times victory did *not* occur.

Reliance on relative frequency information makes sense if relevant events have had an equal number of opportunities to occur or not occur. As Estes (1976a) notes, if the relative frequency of an event happens to correspond to its probability of occurrence (which would have been the case had Estes' subjects witnessed an *equal* number of observation trials for A versus B, and C versus D) then reliance on relative frequency information will lead judges "to make judgments that appear to reflect differences in probabilities of events with great fidelity, but, under slightly different circumstances [for example, when cue frequencies are unequal and frequency of occurrence is unrelated to likelihood of occurrence], the equally efficient operation of the same learning process leads (subjects) to make judgments of likelihoods of events that are widely at variance with the actual probabilities" (p. 51). It is not known how far one can get in our world by deriving predictions of event likelihoods simply from information about relative frequency, but it is clear that the strategy is fraught with dangers.

Estes' research demonstrates that human judges have difficulty with simple comparative proportional estimations. It is therefore not surprising that judges are disinclined to compare conditional probabilities or to process correlation relevant information (see Jenkins and Ward, 1965; Slovic, 1972a; Shweder, 1977c; Ward and Jenkins, 1965). Other, more complex, manipulations of frequency information (such as arriving at an estimate of the proportional *reduction* of error in predicting values on one variable given knowledge of values on a second variable—see Hayes, 1963, p. 608, on the logic of predictive association) are probably beyond the intuitive information processing inclinations of most intelligent adults.

A second hazardous principle for encoding event categories in probability estimation tasks is "verbal association strength" or "conceptual affiliation." Consider the following example:

> Which inference would you endorse?
> M. G. has self-esteem. Therefore, M. G. probably is *not* a leader.
> M. G. has self esteem. Therefore, M. G. probably *is* a leader.

Most informants endorse the second inference. Why? Have they accessed their own frequency beliefs about the conditional probability of being versus not being a leader given that you have self-esteem? If they had, they would have discovered that according to their own beliefs "most people with self-esteem are not leaders" (Shweder, 1977c). Therefore M. G. probably is *not* a leader. What went wrong?

When informants tell us that self-esteem and leadership go together, or draw the inference that someone with self-esteem is likely to be a leader, they are not processing information about the correlation of two variables or the conditional probability of one given the other. What they are doing is judging the extent to which two events co-occur by the extent to which the events affiliate in their minds or have strong verbal associative connections. The conceptual linkages among event classes are relied on to estimate their contingent relationship across persons. Often the result is an "illusory correlation," a consensual estimate of the inductive relationship among events that is not warranted by experience (see Chapman, 1967; Chapman and Chapman, 1967, 1969; D'Andrade, 1974; Shweder, 1975, 1977a; Tversky and Kahneman, 1974). Self-esteem and leadership, for example, go together in our culture's portrait of the ideal leader and in various personifications of that ideal (for example, F. D. R., John Kennedy): They hardly correlate at all across personalities.

There are many ways objects and events conceptually affiliate in our minds. Flavell and Stedman (1961) identify eleven types of "logico-grammatical" relationships for defining the connection between words. These include "similarity" (damp–wet), "supraordination" (animal–fox), "action of" (lion–roar), "action upon" (throw–ball), and "whole–part" (shoe–heel). Casagrande and Hale (1967) needed thirteen types of "semantic relationships" to adequately describe the folk definitions of their Papago Indian informants. These include definition by reference to action sequences, function, common attribute, and

so on. If one is asked how two things "go together" there are obviously many legitimate ways to reply.

In general, conceptual affiliation is a poor index of co-occurence likelihood. Things can be alike without co-occurring. Things can co-occur without being alike. Things can go together in an action sequence without correlating over personalities. For example, blue and green are both alike in hue but that does not entitle us to infer that they co-vary as attributes of "colored things." Similarly, "clings to his mother's apron strings" and "seeks help" are both kinds of "dependent" things to do; they conceptually affiliate in our minds as subordinates of the same category "dependent." Yet this type of conceptual connectedness tells us little about whether those children who "cling to their mother's apron strings" are also the ones who "seek help." The various and diverse ways objects and events relate in our minds is not isomorphic with information about conditional or joint conditional probabilities. A likeness is not a likelihood.

The evidence that we are not intuitively skillful at probability estimation may, at first blush, seem implausible. Many of us seem to assume unwittingly that, over the course of evolutionary history, human survival has, in some way, depended upon our ability to perform high order feats of inductive inferring and probabilistic estimation. Since our species has obviously survived, at least until now, adherence to this assumption makes it necessary to resist or doubt evidence of our limited ability to engage in inductive (and deductive) reasoning or draw correlational lessons from experience.

We would recommend an alternative assumption. In the light of increasing evidence that most intelligent adults are not intuitively inclined to formal operational thinking (see Wason and Johnson-Laird, 1972), we find it preferable to assume that adaptive processes do not require a high level of formal propositional thinking.

Much of the knowledge we possess (and we do possess valid knowledge) is organized in very mundane ways and requires relatively low level inference. Some of this knowledge is episodic and script-like (see Schank and Abelson, 1977); we know what will follow what in a chain of time. Some of it is repetitive and context-specific; we know how so-and-so behaved last time under such-and-such circumstances. There is no necessary reason to assume that mankind's potential for rational adaptive behavior can only be derived in one way, for instance, from some supposed formal operation-like inductive (and deductive) capacity of the individual human mind. In the past, perhaps even today, the context-specific intellectual demands of everyday life may not be very great.

On the other hand, it would be a mistake to conclude that we are always inaccurate in our probability estimations or that we never gain insight into the contingent relationships among events. Estes' point, as we read it, is that it is possible for a God, nature, the environment, or an experimenter, wittingly or unwittingly, to arrange events (for example, by equalizing cue frequencies) so that exclusive reliance on relative frequency information is an accurate index of event likelihoods. There may even be occasions when verbal

association strength and contingency happen to be isomorphic. That is, God, nature, the environment, or an experimenter can help us appear more impressive than we are at probability estimation; even accuracy can be an artifact.

Moreover, Azjen (1977), Tversky and Kahneman (1978), and Ward and Jenkins (1965) have shown that task characteristics can be arranged so that probabilistic and correlational evidence can be easily fit into a pre-existing causal schema or script. Well-scripted evidence is easier to process.

In summary there are two related points to be made about human judgment. The first is that the very same rather limited learning mechanisms and information processing skills can interact with the demands of particular tasks to yield either accurate assessments of reality or massively biased estimations. The second point is that human observers are not infrequently out of accord with the causal and contingent structure of their environment.

Implications for the Study of Implicit Personality Theory

There is a widespread assumption in personality psychology that the lexical items of implicit personality theory (such as honest, responsible, industrious, cooperative, friendly) are labels for scientific categories. Everyday trait terms are often appropriated into academic discourse as though they were inductive summary formulas about the organization of individual differences which encode information about empirical affinities (self-reliant/responsible) and exclusions (gregarious/reserved), and thus have utility in predicting future events and minimizing surprise (Brown, 1965, p. 612; Jackson, Chan, and Stricker, 1979; Passini and Norman, 1966). The systematic distortion hypothesis raises the possibility that the personality categories and classifications of everyday life may not be designed to serve the needs of "man-as-scientist," and should not be analyzed, evaluated, or used as though they were scientific schemes.

Scientists construct classifications and categories for the sake of drawing inductive generalizations ("things that are hot" are "things that hurt;" "people who like parties" are "people who introduce themselves to strangers" — Gilmour, 1937, p. 1040; see also Gilmour, 1951). Given this goal, it follows that the primary criterion for judging the adequacy of a scientific classification is that it be "founded on attributes which have a number of other attributes correlated with them . . ." (Gilmour, 1937, p. 1040). It is important to recognize, especially when studying mundane social cognition, that not all classifications or categories are inductive in intent or designed to serve the inferential purposes of "man-as-scientist."

Many of the conceptual schemes of everyday life are prescriptive (normative) not descriptive in their intent (see Mischel, 1964). They provide "models for" not "models of" reality (Geertz, 1973). They organize the world into categories of events, things, and people for the sake of telling the world how it ought to behave. Since much of human conduct is regulated by rules it would not be surprising if most of the conceptual schemes of everyday life, including implicit personality "theories" (a misnomer?), served a normative

purpose, at least in part. What is a "dog?" Among other things, that's a potentially nutritious animal that one should not eat. *Modern Etiquette in Private and Public* (1872) tells us something about the category "fish." At a dinner party it is to be served directly after the soup, and "you must eat it with a fork, unless silver knives are provided." Science is not all there is to cognition. Mapping correlational structure is not all that the categories of everyday life are about.

There are many non-scientific functions served by our lexicon for personality trait categories. Trait terms are used to inspire conduct ("be courageous"), to proscribe conduct ("stop being so dependent"), and to influence the way others will react. Personality trait labels and trait talk may serve rhetorical, persuasive, and regulatory functions more admirably than they serve predictive or inductive functions.

One can also speculate that personality trait talk serves an important symbolic function in a voluntaristic, individualistic culture like our own. The way a culture allocates individuals to positions in social groups (who is accepted, promoted, retained, confided in, allied with, delegated power and responsibility) expresses or says something about its values, goals, and its views of what is important in man and society. Entrenched Western values make it reprehensible to accept, hire, or promote, either blindly or on the basis of family connections, birthright, color, or caste. Neither a random number table nor a genealogical tree is the right kind of symbol for the image of autonomous man enshrined in Western culture. Selection on the basis of personal character, not chance or connections, is part of our culture's mythic self-conception.

Thus, a priori, there seems to be no reason to assume that everyday personality categories have evolved to serve scientific goals, or to assume that the cultural practices in which trait labels play a part are primarily scientific practices. Everyday trait and type categories (trustworthy, responsible, "the spoiled child") may exist for reasons other than summarizing or encoding nature's regularities. This point is especially important because it implies that the stability and persistence of implicit personality "theories" may be unrelated to the issue of empirical homogeneity and nomological network. One should not expect our everyday trait lexicon to go away simply because it fails to yield valid predictions about what correlates with what across personalities, nor should one doubt that our trait lexicon fails to yield valid predictions simply because it has not gone away.

Summary and a Caution

The systematic distortion hypothesis states that the pattern of correlations among items on memory-based personality instruments tells us more about diverse forms of conceptual affiliation in the minds of raters than about what actually correlates with what across individual differences in conduct, and that these forms of conceptual affiliation can be most easily discovered by simply asking a handful of informants "what is like what?" Examination of the correlational structure of psychiatric ratings using categories from the "Brief Psychiatric Rating Scale" revealed that this rating structure could be reproduced

from semantic judgments about "similarity in meaning." Examination of thirty minutes of videotaped interaction among members of a family revealed that memory-based rating structures parallel pre-existing similarity of meaning structures but do not accurately reflect the correlational structure of actual behavior. Three implications of the systematic distortion hypothesis were discussed:

1. Individual differences are narrowly context-dependent. The neat trait categories and factors so easily retrieved by asking "what is like what" or by examining memory-based rating data are often difficult to induce from actual performance criteria.

2. Normal, intelligent adults are not intuitively skillful at estimating the probability of an event (unless the relative frequency of the event happens to correspond to its likelihood) or the co-occurrence probability of two events (unless the verbal associative bond connecting the two events happens to correspond to their co-occurrence probability).

3. The behavioral categories encoded in our everyday trait lexicon may not have evolved to summarize information about the organization of individual differences, and should not be interpreted as scientific categories.

We conclude this discussion of the systematic distortion hypothesis with a mild caveat: Let the reader beware.

It does not follow from the systematic distortion hypothesis that judges have not tried to faithfully report what they know, or that they know nothing about the ratee, or that any particular item judgment they make is typically erroneous. The point is *not* that raters are characteristically inaccurate. In the videotape study reported earlier, memory-based ratings of *particular* items correlated on the average .30 with measures derived from immediate scorings and in Newcomb's (1929) data rating-scoring correlations for particular items correlated in the .40 to .50 range; presumably, under special conditions favorable to our impressive relative frequency sensors, ratings could correspond even more highly to performance criteria.

The real point is that (1) raters are far from perfect (except under those special circumstances discussed earlier) and when raters do make errors their errors are *systematically biased* errors, not random errors; (2) the typical personality rating situation requires the judge to abstract and summarize a mass of observations from perhaps days, weeks, or months of observation on multiple categories that vary in their base rates and cue frequencies; numerous opportunities for error, and thus systematic bias, occur; (3) if one were to eliminate error and systematic bias from personality data one would not discover neat traits, factors, and dimensions, but rather a complex of context-dependent truths, or alternatively said, that which *is* accurate in personality ratings would not suport a global trait approach to individual differences in conduct.

References

Azjen, I. "Intuitive Theories of Events and the Effects of Base-Rate Information on Prediction." *Journal of Personality and Social Psychology*, 1977, *35*, 303-314.

Block, J. *The Challenge of Response Sets*. New York: Appleton-Century-Crofts, 1965.

Brown, R. *Social Psychology.* New York: Free Press, 1965.
Campbell, D. T., and Fiske, D. W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin,* 1959, *56,* 81–105.
Casagrande, J. B., and Hale, K. L. "Semantic Relationships in Papago Folk-Definitions." In D. Hymes (Ed.), *Studies in Southwestern Ethnolinguistics.* The Hague: Mouton, 1967.
Cattell, R. D. *Description and Measurement of Personality.* New York: World Book, 1946.
Chapman, L. J. "Illusory Correlation in Observational Report." *Journal of Verbal Learning and Verbal Behavior,* 1967, *6,* 151–155.
Chapman, L. J., and Chapman, J. P. "Genesis of Popular but Erroneous Psychodiagnostic Observations." *Journal of Abnormal Psychology,* 1967, *72,* 193–204.
Chapman, L. J., and Chapman, J. P. "Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs." *Journal of Abnormal Psychology,* 1969, *74,* 271–280.
Cronbach, L. J. "Beyond the Two Scientific Disciplines of Scientific Psychology." *American Psychologist,* 1975, *30,* 116–127.
D'Andrade, R. G. "Trait Psychology and Componental Analysis." *American Anthropologist,* 1965, *67,* 215–228.
D'Andrade, R. G. "Cultural Constructions of Reality." In L. Nader and T. W. Maretzki (Eds.), *Cultural Illness and Health.* Washington, D.C.: American Anthropological Association, 1973.
D'Andrade, R. G. "Memory and the Assessment of Behavior." In T. Blalock (Ed.), *Measurement in the Social Sciences.* Chicago: Aldine-Atherton, 1974.
D'Andrade, R. G. "U-Statistical Hierarchical Clustering." *Psychometrica,* 1978, *43,* 59–67.
Douglas, M. *Rules and Meanings.* London: Penguin, 1973.
Ebbesen, E. B., and Allen, R. B. "Further Evidence Concerning Fiske's Question: 'Can Personality Constructs Ever Be Validated?' " Unpublished manuscript, Department of Psychology, University of California, San Diego, 1977.
Einhorn, H. J., and Hogarth, R. M. "Confidence in Judgment: Persistence of the Illusion of Validity." *Psychological Review,* 1978, *85,* 395–416.
Estes, W. K. "The Cognitive Side of Probability Learning." *Psychological Review,* 1976a, *83,* 37–64.
Estes, W. K. "Some Functions of Memory in Probability Learning and Choice Behavior." In G. H. Bower (Ed.), *The Psychology of Learning and Motivation.* Vol. 10. New York: Academic Press, 1976b.
Fiske, D. W. *Strategies for Personality Research: The Observation Versus Interpretation of Behavior.* San Francisco: Jossey-Bass, 1978.
Flavell, J. H., and Stedman, D. J. "A Developmental Study of Judgments of Semantic Similarity." *The Journal of Genetic Psychology,* 1961, *98,* 279–293.
Geertz, C. *The Interpretation of Cultures.* New York: Basic Books, 1973.
Gilmour, J. S. L. "A Taxonomic Problem." *Nature,* 1937, *139,* 1040–1042.
Gilmour, J. S. L. "The Development of Taxonomic Theory Since 1851." *Nature,* 1951, *168,* 400–402.
Hayes, W. L. *Statistics for Psychologists.* New York: Holt, Rinehart and Winston, 1963.
Jackson, D. N., Chan, D. W., and Stricker, L. J. "Implicit Personality Theory: Is It Illusory?" *Journal of Personality,* 1979, *47,* 1–10.
Jenkins, H. M., and Ward, W. C. "Judgment of Contingency Between Responses and Outcomes." *Psychological Monographs,* 1965, *79,* 1–17.
LaForge, R., and Suczek, R. F. "The Interpersonal Dimension of Personality: III, An Interpersonal Checklist." *Journal of Personality,* 1955, *24,* 94–112.
Mischel, T. "Personal Constructs, Rules, and the Logic of Clinical Activity." *Psychological Review,* 1964, *71,* 180–192.
Mischel, W. *Personality and Assessment.* New York: Wiley, 1968.

Mischel, W. "Towards a Cognitive Social Learning Reconceptualization of Personality." *Psychological Review*, 1973, *80*, 252–283.

Modern Etiquette in Private and Public. London: Warne, 1872.

Moos, R. H. "Sources of Variance in Responses to Questionnaires and in Behavior." *Journal of Abnormal Psychology*, 1969, *74*, 405–412.

Mulaik, S. A. "Are Personality Factors Raters' Conceptual Factors?" *Journal of Consulting Psychology*, 1964, *28*, 506–511.

Newcomb, T. M. "The Consistency of Certain Extrovert-Introvert Behavior Patterns in 51 Problem Boys." *Contributions to Education*, 1929, *382*.

Newcomb, T. M. "An Experiment Designed to Test the Validity of a Rating Technique." *Journal of Educational Psychology*, 1931, *32*, 279–289.

Norman, W. T. "Toward an Adequate Taxonomy of Personality Attributes: Replicated Factor Structure in Peer Nomination Personality Ratings." *Journal of Abnormal and Social Psychology*, 1963, *66*, 574–83.

Osgood, C. E. "Interpersonal Verbs and Interpersonal Behavior." In J. L. Cowan (Ed.), *Studies in Thought and Language*. Tucson: University of Arizona Press, 1970.

Overall, J. E., and Gorham, D. R. "The Brief Psychiatric Rating Scale." *Psychological Reports*, 1962, *10*, 799–812.

Overall, J. E., and Hollister, L. E. "Studies of Quantitative Approaches to Psychiatric Classification." In M. M. Katz, J. O. Cole, and W. E. Barton (Eds.), *The Role and Methodology of Classification in Psychiatry and Psychotherapy* (U.S. Public Health Service Publication No. 1584). Washington, D.C.: U.S. Government Printing Office, 1968.

Overall, J. E., Hollister, L. E., and Pichot, R. "Major Psychiatric Disorders: A Four-Dimensional Model." *Archives of General Psychiatry*, 1967, *16*, 146–151.

Passini, F. T., and Norman, W. T. "A Universal Conception of Personality Structure?" *Journal of Personality and Social Psychology*, 1966, *4*, 44–49.

Raush, H. L., Dittmann, A. T., and Taylor, T. J. "Person, Setting, and Change in Social Interaction." *Human Relations*, 1959, *12*, 361–377.

Schank, R., and Abelson, R. *Scripts, Plans, Goals, and Understanding*. Hillsdale, N.J.: Erlbaum, 1977.

Sears, R. R. "Dependency Motivation." In M. R. Jones (Ed.), *Nebraska Symposium on Motivation*. Lincoln: University of Nebraska Press, 1963.

Shweder, R. A. "Semantic Structures and Personality Assessment." Doctoral dissertation, Department of Social Relations, Harvard University. University Microfilms, Ann Arbor, Michigan. Order No. 72-29, 584, 1972.

Shweder, R. A. "How Relevant Is an Individual Difference Theory of Personality?" *Journal of Personality*, 1975, *43*, 455–484.

Shweder, R. A. "Illusory Correlation and the M. M. P. I. Controversy." *Journal of Consulting and Clinical Psychology*, 1977a, *45*, 917–924.

Shweder, R. A. "Illusory Correlation and the M. M. P. I. Controversy: Author's Reply to Some of the Allusions and Elusions in Block's and Edwards' Commentaries." *Journal of Consulting and Clinical Psychology*, 1977b, *45*, 936–940.

Shweder, R. A. "Likeness and Likelihood in Everyday Thought: Magical Thinking in Judgments about Personality." *Current Anthropology*, 1977c, *18*, 637–648.

Shweder, R. A. "Rethinking Culture and Personality Theory, Part I: A Critical Examination of Two Classical Postulates." *Ethos: Journal of the Society for Psychological Anthropology*, 1979a, *7*, 255–278.

Shweder, R. A., and D'Andrade, R. G. "Accurate Reflection or Systematic Distortion? A Reply to Block, Weiss, and Thorne." *Journal of Personality and Social Psychology*, 1979b, *37*, 1075–1084.

Slovic, P. "From Shakespeare to Simon: Speculations—and Some Evidence—About Man's Ability to Process Information." Oregon Research Institute. Research Monograph Vol. 12, No. 12, 1972a.

Slovic, P. "Information Processing, Situation Specificity, and the Generality of Risk-Taking Behavior." *Journal of Personality and Social Psychology*, 1972b, *22*, 128–134.
Smedslund, J. "The Concept of Correlation in Adults." *Scandinavian Journal of Psychology*, 1963, *4*, 165–173.
Smith, G. M. "Usefulness of Peer Ratings of Personality in Educational Research." *Educational and Psychological Measurement*, 1967, *27*, 967–984.
Tversky, A., and Kahneman, D. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology*, 1973, *5*, 207–232.
Tversky, A., and Kahneman, D. "Judgment Under Uncertainty: Heuristics and Biases." *Science*, 1974, *185*, 1124–1131.
Tversky, A., and Kahneman, D. "Causal Schemata in Judgments Under Uncertainty." In M. Fishbein (Ed.), *Progress in Social Psychology*. Hillsdale: Erlbaum, 1978.
Ward, W. C., and Jenkins, H. M. "The Display of Information and the Judgment of Contingency." *Canadian Journal of Psychology*, 1965, *19*, 231–241.
Wason, P. C., and Johnson-Laird, P. N. *Psychology of Reasoning*. London: B. T. Batsford, 1972.
Whiting, B. B., and Whiting, J. W. M. *Children of Six Cultures*. Cambridge, Mass.: Harvard University Press, 1975.
Yarrow, M. R., and Waxler, C. Z. "Dimensions and Correlates of Prosocial Behavior in Young Children." *Child Development*, 1976, *47*, 118–125.

Richard A. Shweder is associate professor of human development in the Committee on Human Development, Department of Behavioral Sciences, University of Chicago.

Roy G. D'Andrade is professor of anthropology in the Department of Anthropology, University of California, San Diego.