

- Shweder, R. A. Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. *Current Anthropology*, 1977, 18, 637-648. (a)
- Shweder, R. A. Illusory correlation and the MMPI controversy. *Journal of Consulting and Clinical Psychology*, 1977, 45, 917-924. (b)
- Shweder, R. A. Illusory correlation and the MMPI controversy: Reply to some of the allusions and elusions in Block's and Edwards' commentaries. *Journal of Consulting and Clinical Psychology*, 1977, 45, 936-940. (c)
- Shweder, R. A. Factors and fictions in person perception: A reply to Lamiell, Foss, and Cavence. *Journal of Personality*, 1980, 48, 54-73.
- Shweder, R. G., & D'Andrade, R. G. Accurate reflection or systematic distortion? A reply to Block, Weiss, & Thorne. *Journal of Personality and Social Psychology*, 1979, 37, 1075-1084.
- Shweder, R. A., & D'Andrade, R. C. The systematic distortion hypothesis. *New Directions for Methodology of Social and Behavioral Science*, 1980, 4, 37-58.
- Stephenson, W. *The study of behavior: Q-technique and its methodology*. Chicago, Ill.: Univ. of Chicago Press, 1953.
- Tyler, L. Toward a workable psychology of individuality. *American Psychologist*, 1959, 14, 75-81.
- Tyler, L. *Individuality: Human possibilities and personal choice in the psychological development of men and women*. San Francisco: Jossey-Bass, 1978.
- Wegner, D. M., & Vallacher, R. R. *Implicit psychology: An introduction to social cognition*. London and New York: Oxford Univ. Press, 1977.
- Wiggins, J. S. *Personality and prediction: Principles of personality assessment*. Reading, Mass.: Addison-Wesley, 1973.
- Wiggins, N. Individual differences in human judgments: A multivariate approach. In L. Rappaport & D. A. Summers (Eds.), *Human judgment and social interaction*. New York: Holt, 1973. Pp. 110-142.
- Zanna, M. P., Olson, J. M., & Fazio, R. H. Attitude-behavior consistency: An individual difference perspective. *Journal of Personality and Social Psychology*, 1980, 38, 432-440.

FACT AND ARTIFACT IN TRAIT PERCEPTION: THE SYSTEMATIC DISTORTION HYPOTHESIS¹

Richard A. Shweder

COMMITTEE ON HUMAN DEVELOPMENT
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS

I. Introduction: The Systematic Distortion Hypothesis	65
II. The Systematic Distortion Hypothesis: Testing the Hypothesis	67
A. The Fit between Conceptual Association Matrices and Rated Behavior Matrices	71
B. The Discrepancy between Conceptual Association Matrices and Actual Behavior Matrices	78
C. The Discrepancy between Rated Behavior Matrices and Actual Behavior Matrices	79
D. Potential Mismisunderstandings: A First Cut	80
III. The Systematic Distortion Hypothesis and Personality Structure	82
A. The Global Trait Model	82
B. The Failure of the Global Trait Model	84
C. Potential Mismisunderstandings: A Second Cut	85
IV. The Systematic Distortion Hypothesis and Implicit Personality Theory	88
A. The Uses of a Trait Lexicon: Scientific vs Nonscientific Functions	89
B. Potential Mismisunderstandings: A Third Cut	92
V. Conclusion: From Likeness to Likelihood	95
References	96

I. Introduction: The Systematic Distortion Hypothesis

The systematic distortion hypothesis states that our beliefs about personality structure tend to be inaccurate with respect to how attitudes, affects, and behaviors covary (see D'Andrade, 1965, 1973, 1974; Shweder, 1972, 1973, 1975, 1977a, 1977b, 1977c, 1979a, 1980a; Shweder & D'Andrade, 1979, 1980). The

¹This article is a revision and synthesis of sections of two unpublished manuscripts entitled "Fact and Artifact in Personality Assessment: The Influence of Conceptual Schemata on Individual Difference Judgments," and "Attributional Illusions in Psychological Theory: The Firsaken Quest for an Individual Difference Theory of Personality." The article also draws on formulations and materials in Shweder (1972, 1975, 1979a), and Shweder and D'Andrade (1979, 1980). Responsibility for the views set forth here, however, is entirely my own.

main idea is that, under the memory conditions characteristic of most lay and scientific personality assessments, judges either infer correlational structure from a general model of conceptual association, or find conceptually associated memory items easier to retrieve. In other words, *inferences* about personality contain a systematic bias in that propositions about "what is like what" are substituted for propositions about what is likely, and *memory* for personality relevant events contains a systematic bias in that attitudes, affects, and behaviors that are conceptually associated (e.g., "aggression" and "dominance," "disagrees" and "criticizes") are recalled as if they covaried. The Chapmans have labeled this effect "illusory correlation" (Chapman, 1967; Chapman & Chapman, 1967, 1969).

The systematic distortion hypothesis concerns the limitations of our current knowledge of personality syndromes. Personality syndrome is interpreted here as a package of correlated affective, attitudinal, and behavioral characteristics. Personality syndromes are sometimes referred to as global traits, underlying factors, or general dimensions. "Introversion" is an example of a postulated syndrome; the syndrome is postulated with the expectation that the correlational structure of individual differences across comparable contexts will reveal a discriminated cluster of positively correlated attributes, including "prefers to work alone," "slow to make friends," "daydreams at meetings." The personality psychology literature is rich in postulated syndromes: the "dependent" child (seeks help, seeks reassurance, seeks attention, seeks physical proximity); the "paranoid" adult (hostile, suspicious, grandiose); the "permissive" mother (feeds her children on demand, refrains from physical punishment, accepts nudity, tolerates masturbation). Hundreds of postulated syndromes extant in the literature could be cited: "ego-strength," "Machiavellianism," "anxiety," "psychoticism," "ascendancy," "altruism."

One implication of the systematic distortion hypothesis is that one cannot trust evidence of personality syndromes derived from memory-based assessment procedures (inventories, rating forms, questionnaire interviews). And, since most evidence in support of proposed personality syndromes is of this memory-based type, a second implication is that the very idea that people have global traits consisting of covarying behaviors may be illusory, the product of a widespread human tendency to rely on similarity and conceptual proximity for estimating cooccurrence probability (Tversky & Kahneman, 1974).

An explicit, contemporary version of the systematic distortion hypothesis was first advanced by D'Andrade in 1965. An earlier, exemplary formulation can be found in Newcomb (1929, 1931). A cognate formulation can be found in Mulaik (1964). Since 1965 the hypothesis has been sympathetically received in some quarters (Mischel, 1968, 1973; Berman & Kenny, 1976; Fiske, 1978; Ibbesen & Allen, 1977, 1979; Cooper, 1981; Gara & Rosenberg, 1981; Mirrels, n.d.; Nisbett, 1980; Berman, 1982) and criticized in others (Block, 1977; Block, Weiss, & Thorne, 1979; Epstein, 1979; Jackson, Chan, & Stricker, 1979; Lamiell, Foss,

& Cavence, 1980). Replies to criticism can be found in Shweder (1977b, 1979a, 1980a) and Shweder and D'Andrade (1979).

Unfortunately, for the sake of scientific progress, the systematic distortion hypothesis seems to be easily misunderstood, apparently lending itself to hyperbolic formulations that detract from rational debate and are unlikely to produce pertinent empirical investigations. Indeed, anyone so unfortunate as to read only the literature critical of the systematic distortion hypothesis might well come away with the misbegotten notion that the hypothesis denies the existence of personality, claims that raters make ratings ignorant of rates, or hypothesizes a social process in which people interact with each other in a linguistic fantasy world!

The systematic distortion hypothesis also seems to generate a good deal of resistance. Global trait concepts (dominant, aggressive, friendly, dependent, etc.) are deeply entrenched in most natural languages and in lay personality theory (see Shweder, 1972; White, 1980; Shweder & Bourne, 1981, for cross-cultural evidence). Many readers probably subscribe to the widespread assumption that natural language categories and implicit personality theory would not persist unless they were more or less valid. Indeed, there are even some contemporary theories of category formation which argue that lay categories are basically encodings of "real world" correlational structures (Rosch, 1975; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Thus, any hypothesis that would deny that our beliefs about correlational structure are more or less accurate summaries of the actual "relative frequencies of joint occurrences of various personality attributes and behavioral dispositions in other persons" (Passini & Norman, 1966, p. 47) has some explaining to do.

The aim of this article is to eliminate misunderstandings and, hopefully, to overcome resistance (also see Shweder, 1979a, 1980; Shweder & D'Andrade, 1979). The systematic distortion hypothesis shall be described and a corpus of studies supporting the hypothesis shall be reviewed. The implications of the hypothesis shall be enumerated with special reference to personality structure and implicit personality theory. Throughout the article pauses are taken to anticipate possible misunderstandings.

II. The Systematic Distortion Hypothesis: Testing the Hypothesis

The systematic distortion hypothesis asserts that judges on personality inventories, interpersonal checklists, and questionnaire interviews unwittingly substitute a preexisting model of conceptual association for a description of correlational structure. Thus, the correlational structure of ratings replicates preexisting beliefs about what is like what with little sensitivity to the correlational structure of actual behavior.

One way to test the systematic distortion hypothesis is to collect conceptual

association judgments, memory-based ratings, and observational evidence on an equivalent set of attitudinal, affective, or behavioral variables. By a conceptual association judgment I mean a direct judgment of "similarity in meaning" between pairs of descriptive terms or phrases, a derived measure of conceptual proximity based on common associates in a free association task, or the relative number of predications common to the pairs of terms or phrases being judged (see, e.g., Szalay & Deese, 1978). A summary of the conceptual association judgments among all possible pairs of variables in a set of attitudinal, affective, or behavioral variables shall be referred to as a "conceptual association matrix." By observational evidence I mean a reliable, reasonably objective, "on-line" record made at the time of observation. Such records will be referred to as "immediate scorings," and a summary of the intercorrelations in immediate scorings among all possible pairs of variables in a set of attitudinal, affective, or behavioral variables shall be referred to as an "actual behavior matrix." By memory-based ratings I mean the type of judgment characteristic of personality assessment instruments such as the MMPI, the California Q-Sort, or the Brief Psychiatric Rating Scale. Judges are asked to abstract and summarize their previous observations of themselves or others on a set of attitudinal, affective, or behavioral variables. The intercorrelations in memory-based ratings among all possible pairs of variables in a set of attitudinal, affective, or behavioral variables shall be referred to as a "rated behavior matrix."

The systematic distortion hypothesis predicts that the correlational structure of variables in rated behavior matrices is unlike that of equivalent variables in actual behavior matrices, yet replicates the patterning of conceptual association judgments among those same variables in conceptual association matrices. In other words, what correlates with what in memory-based ratings tells us more about preexisting ideas of what is like what than about what correlates with what in actual behavior. This predicted pattern of results is illustrated in Fig. 1.

The data presented in Fig. 1 are from a videotape study conducted by R. G. D'Andrade and reported in Shweder and D'Andrade (1980). The subset of data presented in Fig. 1 is selected for illustrative purposes only. In the study reported in Shweder and D'Andrade, 30 minutes of videotaped interaction among four members of a white, middle-class California family were analyzed using 11 categories of interpersonal behavior (6 of these categories are selected for illustration in Fig. 1).

Three observers conducted on-line immediate scorings of behavioral acts using the 11 categories. The mean reliability coefficient across all 11 categories for percentage of act across actors was .75. These immediate scorings were used to construct an actual behavior matrix for all pairs of 11 categories.

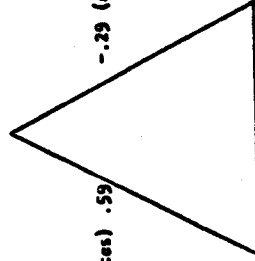
Twenty raters viewed the 30-minute videotape and then gave summary ratings of each actor on each category using a 7-point scale. Raters were asked "how much does [so-and-so] do the following [for example, 'criticize']?" These

Conceptual association matrix
(similarity of meaning judgments averaged across 10 judges)

	Ad	In	Su	Qu	Cr	Di
Advise	76	88		-04	-08	-36
Inform		64		-12	-12	-28
Suggest				-16	-28	16
Question					44	72
Criticize						44
Disagree						

(r_s between matrices) .59

- .29 (r_s between matrices)



.22 (r_s between matrices)

Rated behavior matrix
(mean r coefficients averaged over 20 memory-based raters)

	Ad	In	Su	Qu	Cr	Di
Advise	42	51		24	10	05
Inform		37		14	00	-10
Suggest				11	13	-02
Question					12	-01
Criticize						59
Disagree						

Actual behavior matrix
(r coefficients for behavior percentage averaged across immediate on-line scorings of three reliable observers)

	Ad	In	Su	Qu	Cr	Di
Advise	00	00		00	67	33
Inform		-33		33	-33	00
Suggest				33	33	-67
Question					-33	-67
Criticize						00
Disagree						

FIG. 1. The systematic distortion hypothesis. Degrees of correspondence (r_s) between correlational structures derived from similarity of meaning judgments (conceptual association matrix), memory-based ratings (rated behavior matrix), and immediate scorings (actual behavior matrix). Subset of data from videotape study (Shweder & D'Andrade, 1980) selected to illustrate typical pattern of findings.

memory-based ratings were used to construct a rated behavior matrix for all pairs of 11 categories.

Ten judges made similarity of meaning judgments for all pairs of lexical labels for the 11 interpersonal behavior categories. Judgments were made on a scale running from +100 ("identical in meaning") to -100 ("completely opposite in meaning"). These similarity of meaning judgments were used to construct a conceptual association matrix for all pairs of 11 categories.

The conceptual association matrix, the rated behavior matrix, and the actual behavior matrix were compared directly with each other by correlating over the parallel cells of each pair of matrices. This is one way of asking whether the correlational structure of each matrix reveals the same information about what goes with what across individual differences in conduct.

Using all 11 categories the correspondence between the conceptual association matrix and the rated behavior matrix was substantial ($r = .75$), while that between the rated behavior matrix and the actual behavior matrix ($r = .22$) and the actual behavior matrix and the conceptual association matrix ($r = .00$) was either weak or nonexistent. For the subset of six categories selected for illustrative purposes in Fig. 1, the measures of intermatrix correspondence are .59, .22, and -.29, respectively. Across seven different tests of intermatrix correspondence (D'Andrade, 1974; Shweder, 1975, 1977a; Shweder & D'Andrade, 1980), the mean intermatrix correlation between conceptual association matrices and rated behavior matrices has been .75, the mean intermatrix correlation between rated behavior matrices and actual behavior matrices has been .25, and the mean intermatrix correlation between actual behavior matrices and conceptual association matrices has been .26.

What has been discovered is the pattern of results predicted by the systematic distortion hypothesis: (a) Conceptual associations among event descriptors (e.g., "disagrees" and "criticizes") do not parallel the correlational structure of actual behavior (e.g., "disagrees" and "criticizes" are uncorrelated; see Fig. 1), at least not across individual differences in conduct as immediately scored; (b) the correlational structure of rated behavior does not parallel the correlational structure of actual behavior; and (c) the correlational structure of rated behavior does parallel our lay model of conceptual associations among event descriptors.

The overall pattern of results lends itself to the interpretation that memory-based reports about the correlational structure of individual differences are really reflections of a preexisting lay model of conceptual association, and that conceptual association is a poor index of concurrence likelihood in the real world.

A noteworthy distinguishing feature of actual behavior matrices vs. rated behavior matrices and conceptual association matrices is the relative absence of neat syndromes, global traits, or simple main effects in actual behavior matrices, and the striking presence of these broad underlying factors in conceptual association and rated behavior matrices. Thus, in Fig. 1, for example, the conceptual association matrix displays a neat differentiation of two "styles" of interpersonal

behavior, *viz.* advise, inform, suggest vs. question, criticize, disagree. That clean partitioning of behavioral traits blurs slightly in the rated behavior matrix but is still detectable, while in the actual behavior matrix one is hard-pressed to retrieve any underlying general factors. It is as if the celestial mind of man conceives of pure global types which the mundane world of behavior refuses to instantiate. As the reader can see, there are three pieces to the pattern of evidence required to support the systematic distortion hypothesis. Each shall now be considered in turn.

A. THE FIT BETWEEN CONCEPTUAL ASSOCIATION MATRICES AND RATED BEHAVIOR MATRICES

Personality researchers share with other scientists the image of the world as "a scene of recurrent kinds of events and changes which exemplify certain regular connections" (Hart, 1961, p. 184). In keeping with this image, a major goal for personality researchers has been to devise a parsimonious set of syndromes, traits, factors, or dimensions (e.g., egoistic vs. altruistic, extroverted vs. introverted) for predicting individual differences in one context from knowledge of individual differences in another one. In other words, personality researchers *qua* taxonomists have sought to construct categories which summarize regularities in behavior and enable them to make relatively valid inductive generalizations (e.g., children who "offer help" are children who "make suggestions," "people who like parties" are "people who introduce themselves to strangers"). In the quest for an adequate personality classification, correlational structure has been a central concern (see Gilmour, 1937, 1951; Sokal & Sneath, 1963).

Perhaps the most fundamental criterion for evaluating a scientific classification is that it be "founded on attributes which have a number of other attributes correlated with them" (Gilmour, 1937, p. 1040), or, said alternatively, whatever the scheme of classification there must be a number of true and relevant statements which can be made regarding its constituent categories (e.g., "dependent" children are more likely than other children to seek help from peers, seek attention from adults, cling to their mothers' apron strings).

Correlational structure has been a central concern of personality researchers (e.g., Cattell, 1946; Sears, Maccoby, & Levin, 1957; Norman, 1963; Block, 1965; Overall, Hollister, & Pichot, 1967; Bales, 1970). Most personality taxonomies are the product of a data reduction process in which a relatively large number of individual difference attributes (e.g., self-reliance, responsibility, assertiveness) are sorted into a relatively small number of underlying syndromes, traits, or factors (e.g., character strength). Typically, trait equivalence and difference are determined by the extent to which sets of attributes covary, or said alternatively, by the extent to which knowledge of one attribute (e.g., he's "self-reliant") enables one to make valid predictions about other attributes (e.g., he's probably "responsible," "assertive"), and vice versa.

TABLE I

BALES' 26-ITEM RATING FORM FOR PERSONALITY AND INTERPERSONAL BEHAVIOR*

1. Respectful
2. Admired
3. Dominating
4. Speaks like an autocratic authority
5. Devalues himself
6. Makes others feel he admires them
7. Especially addressed when others have serious opinions about which they want confirmation
8. Stands for the most conservative ideas and beliefs of the group
9. Rejects religious belief generally
10. Liked
11. Feels that others are generally too conforming to conventional social expectations
12. Always tries to speak objectively
13. Thinks of himself as entertaining
14. Warm and personal
15. Accepts failure and withdrawal for himself
16. Identifies himself with the underprivileged
17. Demands pleasure and gratification
18. Believes that equality and humanitarian concerns for others are very important
19. Introverted
20. Believes it is necessary to sacrifice the self for higher values
21. Assumes responsibility for task leadership
22. Receives a lot of interaction from others
23. Valuable for a logical task
24. Feels that his individual independence is very important
25. Personally involved in the group
26. Passively withholds cooperation

*Adapted from Bales (1970).

dimensions, on the one hand, and the three corresponding conceptual association dimensions, on the other. Conceptual association judgments yield patterns of interitem proximity similar to the correlational structure of rated behavior.

Most personality classifications derived from memory-based assessment procedures can be reproduced from conceptual association judgments. To date the following memory-based personality classifications have been successfully retrieved using conceptual association techniques.

1. Factor-analytic classification of personality adjectives, as given in Norman (1963). See D'Andrade (1965) and Mulaik (1964).
2. Leary grid organization of interpersonal behavior, as given in LaForge and Sucek (1955). See D'Andrade (1965).
3. Factor-analytic classification of personality and interpersonal behavior, as given in Bales (1970). See Shweder (1972, 1975).

It is noteworthy that "impressive" personality taxonomies (e.g., LaForge & Sucek, 1955; Norman, 1963; Block, 1965; Smith, 1967; Wiggins, 1978) have been constructed almost entirely out of patterns of response covariation on memory-based checklists, inventories, and questionnaires. By an impressive taxonomy I mean one that is both parsimonious (e.g., a 30-item checklist is reduced to three underlying global traits) and valid (i.e., from knowledge of a subject's response to one item on the checklist one can draw inferences about how the subject responded to other items on it, and those inferences, using probabilistic criteria, turn out to be true). It is also noteworthy that most of the memory-based taxonomies extant in the personality literature can be reproduced by asking a handful of respondents to judge the degree of conceptual association (e.g., "similarity of meaning") of the items on the checklist, inventory, or questionnaire.

The underlying taxonomic structure of most memory-based personality classifications can be replicated from conceptual association (or similarity of meaning) judgments without having to rate anyone's personality. Table II illustrates this point with special reference to Bales' three-factor classification of personality and interpersonal behavior in small groups.

Bales' three-factor classification ("power," "likeability," and "task-orientation") was first induced from the correlational structure of hundreds of individual difference variables on 60 subjects who participated in 12 five-person groups. Most of Bales' measures were memory-based, including three personality inventories (MMPI, Cattell's 16 PF test, and Thurstone's temperament test) and ratings by peers and outside observers. Bales then devised a 26-item inter-personal rating form to diagnose personality and classify group members in his three-dimensional scheme. The 26 items are listed in Table I.

Shweder (1975) analyzed the correlational structure of a rated behavior matrix for the 26 items in Table I and compared it with the correlational structure of a conceptual association matrix for the same items. The rated behavior matrix was derived from the peer ratings of members of a self-analytic group, and was scaled in three-dimensional space (MDSCALE) (see Table II). The conceptual association matrix was derived from sorting task data where judges were asked to place items "similar in meaning" together in the same pile. Pairs of items appearing in the same pile across judges received higher conceptual association scores. The resulting conceptual association matrix was also scaled in three-dimensional space (MDSCALE) (see Table II). The rated behavior matrix was compared to the conceptual association matrix by determining the extent to which the two scaling solutions yielded identical spatial coordinates for three dimensions. This comparison is also shown in Table II.

Table II lists the three-dimensional spatial coordinates for the 26 items (Table I) in both rated behavior and the preexisting conceptual association model of judges. The two classifications of the items are quite similar, as indicated by rank-order correlations (r 's) of .86, .88, and .70 between the three rated behavior

TABLE II
BALES' THREE DIMENSIONS OF PERSONALITY AND INTERPERSONAL BEHAVIOR*

Item number	Power		Likability		Task-orientation	
	Conceptual dimension	Rating dimension	Conceptual dimension	Rating dimension	Conceptual dimension	Rating dimension
1	-.38	.01	-.89	-.99	-.12	-.44
2	.18	.34	.74	.71	.27	.30
3	.71	.77	-.82	-.54	.06	.13
4	.63	.65	-1.00	-.71	-.05	.21
5	-1.13	-.99	-.06	.12	-.18	-.43
6	-.18	-.15	.89	1.06	-.00	-.00
7	.62	.66	.41	.31	.49	.36
8	.43	-.34	-.68	-.78	.29	.67
9	-.09	.03	-.44	-.12	.30	-.96
10	.65	-.01	.89	1.00	-.27	.20
11	-.23	.16	-.53	-.53	-.66	-.73
12	.00	-.02	.21	-.02	.95	1.00
13	.65	.64	.00	-.13	-.46	-.65
14	.38	-.10	1.04	1.03	-.36	-.18
15	-1.22	-1.14	-.09	-.46	.16	-.48
16	-.71	-.72	-.22	.36	-.71	-.41
17	.45	.51	-.55	-.78	-.48	-.43
18	-.29	-.41	.61	.80	-.38	-.01
19	-1.12	-1.28	-.26	-.15	.38	.17
20	-.70	-.46	-.09	.42	-.48	.49
21	.79	.75	.06	.02	.20	.44
22	.73	.79	.85	-.03	.16	.09
23	.28	.47	.45	.42	.92	.66
24	-.05	.34	-.83	-.75	.18	.24
25	.34	.40	1.11	.64	.00	-.05
26	-.76	.90	.78	.90	-.19	-.19

* Conceptual association and memory-based rating dimensions compared. MIMSCAL: spatial coordinates.

4. Factor-analytic classification of maternal personality, as given in Sears *et al.* (1957). See Shweder (1975).

5. Correlation matrices for Bales Interaction Process Analysis categories used as rating scales, as given in Borgatta, Cottrell, and Mann (1958) and in Mann (1959). See D'Andrade (1974).

6. Partial correlation matrix for observers rating extroversion-introversion in boys, as given in Newcomb (1929). See Shweder (1975, 1977a).

7. The Alpha factor of the MMPI, as given in Block (1965). See Shweder (1977a, 1977b).

8. The common-factor structure of the Murray Needs from five personality tests as given in Fiske (1973) and in Huda and Hamilton (1976). See Ebbesen and Allen (1977).

9. Syndrome clusters from the Brief Psychiatric Rating Scales, as given in Overall *et al.* (1967). See Shweder and D'Andrade (1980).

10. Trait clusters of personality adjectives, as given in Rosenberg, Nelson, and Vivekananthan (1968). See Gara and Rosenberg (1981).

11. Factor-analytic classification of California Q-set variables, as given in Block *et al.* (1979). See Block *et al.* (1979).

The results of all these studies are remarkably uniform: the correlational structure of rated behavior matrices and the personality taxonomies retrieved from rated behavior matrices closely resemble a model of conceptual associations preexisting in the minds of raters.

The fit between conceptual association matrices and rated behavior matrices is a necessary but not sufficient piece of the evidence required to support the systematic distortion hypothesis. It is necessary because the hypothesis claims that the correlational structure of memory-based ratings is a reflection of judges' preexisting models of conceptual association but it is not sufficient because the fit between conceptual association matrices and rated behavior matrices is also consistent with an alternative "accurate reflection" hypothesis. As noted in Shweder and D'Andrade (1979, p. 1076), this alternative hypothesis "asserts that ordinary folk learn or develop 'implicit personality theories' that summarize and preserve the empirical covariation of behavior across individual differences in conduct. According to this lacunae reflection hypothesis, people use empirically valid implicit personality theories in making conceptual similarity judgments, thereby accurately reporting the intercorrelations of behaviors." Advocates of the accurate reflection hypothesis include Passini and Norman (1966), Block *et al.* (1979), and Jackson *et al.* (1979).

What are conceptual association matrices all about? What are subjects telling us when they judge two personality descriptors (e.g., aggressive/dominant) similar in meaning? According to the systematic distortion hypothesis (Shweder & D'Andrade, 1979, p. 1081), similarity of meaning judgments are based on the relative number of predications common to the two terms being judged. There are many types of true things one can say about particular pairs of terms, for example, that both behaviors referred to by the terms make one (e.g.) angry, that the referred-to behaviors follow each other in a sequence, or share common characteristics. According to the systematic distortion hypothesis, similarity of

meaning judgments vary directly with the ease with which subjects bring these predications to mind, and most of the truths which subjects bring to mind in making similarity of meaning judgments reveal little about the likelihood that the behaviors covary across individual differences in conduct. In contrast, according to the accurate reflection hypothesis, similarity of meaning judgments are more or less accurate probabilistic summaries of the actual correlational structure of individual differences.

Are similarity of meaning judgments reports about non-correlational conceptual associations (systematic distortion hypothesis) or are they accurate summaries of correlational structure (accurate reflection hypothesis)? There are at least three methods for choosing between these two interpretations of conceptual association matrices.

One method is to elicit detailed introspective accounts of the reasoning that produces a similarity of meaning judgment. Ask subjects what they were thinking of when they judged (e.g.) "self-esteem" and "leadership" or "aggression" and "dominance" more similar in meaning than "pretentious" and "reclusive." Research by Shweder (1977c) suggests that ordinary folk have great difficulty with correlational reasoning (also see Smedslund, 1963; Ward & Jenkins, 1965; Jenkins & Ward, 1965; Crocker, 1981), and rely instead on semantic relationships other than correlation in making a similarity of meaning judgment (also see Shweder, 1972.)

Some pairs of terms, for example, "hostile" and "warm," are judged dissimilar in meaning either because the judge cannot imagine applying both terms to the same *act*, or because the two terms evoke opposed connotations (re: Osgood, Suci, & Tannenbaum, 1957). Those two types of semantic exclusions, of course, tell us nothing about whether people who tend to engage in more "hostile" acts than others are also less likely than others to engage in "warm" acts.

Other pairs of terms, for example, "aggression" and "dominance," are judged similar in meaning because they are sequentially linked segments in a common "script" (re: Abelson, 1976), viz. "fighting to get ahead in a hierarchy." This type of conceptual association, a scripted sequential dependency, again tells us little about individual differences, that is, whether people who exercise their authority more than others also tend to behave more aggressively.

Still other pairs of terms, for example, "self-esteem" and "leadership," are judged similar in meaning because they coexist as elements in an idealized behavioral type. Indeed many subjects estimate that there are far more "people with self-esteem" in the population than there are leaders, a base-rate estimate consistent with the view that most people with self-esteem are *not* leaders. However, subjects do not think of base-rates (see Lyon & Slovic, 1976) when they judge "self-esteem" and "leadership" similar in meaning. Instead, they think of Franklin Delano Roosevelt, John Kennedy, or some other embodiment of our culture's portrait of an ideal leader.

A good deal more work must be done on the diverse forms of conceptual association underlying similarity of meaning judgments. Two coding schemes for classifying types of conceptual associations have been proposed by Flavell and Stedman (1961) and Casagrande and Hale (1967) (see also D'Andrade, 1974). The very preliminary work reported in Shweder (1972, 1977c) suggests that similarity of meaning judgments have much to do with sequential scripting, shared denotations and connotations, shared effects, and part-part and part-whole relationships, and rather little to do with a process of correlational reasoning across individual differences in conduct.

Free associational methods (e.g., Szalay & Deese, 1978) provide a second way to choose between systematic distortion and accurate reflection interpretations of similarity of meaning judgments. The systematic distortion hypothesis predicts that direct pairwise similarity of meaning judgments should be reproducible from free associational data. That is, pairs of personality terms judged more similar in meaning should display more common associates on free associational tasks, and measures of conceptual proximity derived from free associational data should correlate with scaled judgments of likeness in meaning. The accurate reflection hypothesis makes no such prediction. Indeed, if similarity of meaning judgments are merely accurate reports of the intercorrelations among behaviors, free associational data should tell us little about what is like what in the minds of subjects.

No one has yet tried to replicate direct pairwise similarity of meaning judgments for personality terms using free associational techniques. Successful replications have been performed with terms from other domains (Szalay & Deese, 1978, pp. 116-120).

Gara and Rosenberg (1981), however, have elicited semantic feature descriptions for each of 20 personality adjectives. Subjects were asked, "What would lead you to label a person _____?" Two conceptual association matrices were derived using two different measures of interitem feature sharing, and these "shared feature" based conceptual association matrices were compared to a conceptual association matrix derived from direct similarity of meaning judgments. The two intermatrix correlations were .61 and .72. D'Andrade (1974) and Shweder (1977c) have argued that conceptual association matrices reflect a variety of types of conceptual relationships, including shared semantic features. In line with this observation, Gara and Rosenberg's (1981) work suggests that feature sharing is an important part, but only one part, of the story. For other techniques for deriving conceptual association matrices see Sticffire, Reich, and McClaren-Sticffire (1971) and D'Andrade, Quinn, Nerlove, and Romney (1972).

There is a third way to choose between systematic distortion vs accurate reflection interpretations of similarity of meaning judgments. According to the systematic distortion hypothesis, conceptual associations (shared features, scripted sequential dependencies, common effects, part-part relationships, etc.)

are poor indices of covariation relationships across individual differences in conduct. In contrast, the accurate reflection hypothesis holds that conceptual association matrices are reflections of implicit personality theories that summarize and preserve the empirical correlational structure of behaviors across personalities. To choose between the two hypotheses one need only compare, using the same items, the interitem similarity structure of conceptual association matrices with the interitem correlational structure of actual behavior matrices. To date comparisons with seven data sets have been performed (D'Andrade, 1974; Shweder, 1975, 1977a; Shweder & D'Andrade, 1980). For the most part, these comparisons support the systematic distortion hypothesis. Conceptual association matrices are weakly related to actual behavior matrices.

B. THE DISCREPANCY BETWEEN CONCEPTUAL ASSOCIATION MATRICES AND ACTUAL BEHAVIORAL MATRICES

There are relatively few behavior observational studies in the personality literature, probably because memory-based rating techniques (inventories, questionnaires, etc.) are so much easier to administer, score, and analyze. There are also relatively few multimethod studies in the personality literature, despite Campbell and Fiske's (1959) now "ancient" warning of the hazards of single method research. To test the systematic distortion hypothesis adequately, it is necessary to compare the correlational structure of equivalent sets of personality variables across conceptual association matrices, rated behavior matrices, and actual behavior matrices. Between D'Andrade (1974), Shweder (1975, 1977a), and Shweder and D'Andrade (1980), seven relevant data sets have been either located or generated. Two data sets each come from Borgatta *et al.* (1958), Mann (1959), and Newcomb (1929). One data set comes from Shweder and D'Andrade (1980).

All these data sets are multimethod sets, utilizing both behavior observational techniques and memory-based ratings. The behavior observational evidence varies from on-line videotape codings (Shweder & D'Andrade, 1980) to immediate scorings of the stream of behavior (Mann, 1959) and "daily records" (Newcomb, 1929) where observers were instructed to note the presence or absence of a comb, 1929) where observers were instructed to note the presence or absence of a type of behavior as soon after its occurrence as possible. For the Shweder and D'Andrade (1980) data, the interscorer reliability coefficients for 11 categories averaged .75. For the Newcomb data (1929), the odd- vs even-day reliability coefficient for 26 categories averaged .78. For the Mann (1959) data, a single observer did the scoring alone, but had been precalibrated to an interscorer reliability criterion of .90 on a majority of categories. Reliabilities in the Borgatta *et al.* (1958) study are more difficult to assess with precision on the basis of the published data (see Shweder & D'Andrade, 1979). Hopefully, this article will entice some readers to generate new multimethod data sets appropriate for testing the systematic distortion hypothesis.

As noted earlier, the accurate reflection hypothesis advocated by Passini and Norman (1966) and Jackson *et al.* (1979) has not fared well in direct comparisons of conceptual association matrices with actual behavior matrices. Across seven tests the average correlation between conceptual association matrices and actual behavior matrices is .26. The range is from -.05 to .51. With regard to the Newcomb data, where the accurate reflection hypothesis has its best showing, there is an average intermatrix correlation of .38 between the conceptual association matrices of each of 10 judges, on the one hand, and the actual behavior matrix of Newcomb's observers (Shweder, 1977a). Yet even this modest degree of fit should probably be discounted since the "immediate scorings" in the Newcomb data were not always "on-line," and were probably themselves subject to modest degrees of systematic distortion in the direction of preexisting conceptual associations.

Overall, similarity of meaning judgments do not seem to reflect empirically valid implicit personality theories; our preexisting models of what is like what do not seem to parallel closely the actual correlational structure of individual differences in conduct. "Aggression" and "dominance" or "criticism" and "disagreement" (e.g.) may associate in our minds but not in the organization of personality.

C. THE DISCREPANCY BETWEEN RATED BEHAVIOR MATRICES AND ACTUAL BEHAVIOR MATRICES

The systematic distortion hypothesis states that the broad syndromes, global traits, underlying factors, and general dimensions retrieved from memory-based rating data (inventories, questionnaires, etc.) tell us more about preexisting, noncorrelational patterns of conceptual association (semantic feature sharing, scripted sequential dependencies, common effects, etc.) than about what correlates with what across individual differences in conduct. In Section II, A I briefly reviewed the ever-growing body of evidence that rated behavior matrices replicate preexisting associative connections among personality descriptors. In Section II, B I noted that what evidence does exist suggests that conceptual associations among personality descriptors are only weakly related to the actual correlational structure of individual differences. The obvious implication is that rated behavior correlations also tell us little about what correlates with what in the structure of personality. Indeed, across the seven multimethod data sets mentioned earlier, the average correlation between rated behavior matrices and actual behavior matrices was .25 (D'Andrade, 1974; Shweder, 1975, 1977a; Shweder & D'Andrade, 1980).

As we have seen, the systematic distortion hypothesis requires evidence on the relationship between three types of data matrices. The ideal way to test the hypothesis is to have raters observe behavior in which what actually correlates with what across raters is unrelated, or opposed, to what associates with what in

the preexisting conceptual association scheme of raters. Will the interitem correlational structure of the (subsequently rated) rated behavior matrix replicate the interitem proximity structure of the preexisting conceptual association matrix, as predicted by the systematic distortion hypothesis, or will it correspond to the interitem correlational structure of the actual behavior matrix?

One such test can be found in Shweder's (1975) reanalysis of Newcomb's (1929) study of extroversion and introversion in the behavior of boys at summer camp. In one part of the reanalysis, Shweder (1975, pp. 464-466) deliberately selected 33 pairs of items (e.g., "speaks with confidence of his own abilities" - "spends more than an hour a day alone") to ensure that conceptual association scores would be radically inconsistent with actual behavior correlations. Each of the 33 item pairs either associated in the minds of judges and was not positively correlated in actual behavior, or did not associate in the minds of judges and did positively correlate in actual behavior. In effect, over the 33 pairs of items, the degree of correspondence between what correlated with what in actual behavior and what associated with what in the minds of raters was set at $-.36$. The interitem proximity structure of preexisting conceptual associations was in radical conflict with the interitem correlational structure of observed individual differences. Under these manipulated circumstances, what happened to the interitem correlational structure of memory-based ratings? Memory-based ratings of what correlated with what replicated the interitem proximity structure of preexisting conceptual associations ($r = .84$) and were insensitive to the interitem correlational structure of observed behavior ($r = -.27$). Similar results have been reported by Chapman and Chapman (1967, 1969).

In the study just described, Shweder preset, at $r = -.36$, the discrepancy between interitem conceptual associations and interitem actual behavior correlations. Under naturally occurring conditions, the fit between conceptual association matrices and actual behavior matrices seems to vary between approximately $.00$ and $.50$, and the validity of rated behavior matrices varies accordingly (D'Andrade, 1974; Shweder, 1975, 1977a; Shweder & D'Andrade, 1980).

To summarize, rated behavior matrices tell us more about diverse noncorrelational forms of conceptual association in the minds of raters than about what actually correlates with what across personalities, and these conceptual association clusters can be most easily discovered by simply asking a handful of subjects "what is like what," without any reference to anyone's behavior.

Does that mean that raters are blind, that we live in a linguistic fantasy world, or that personality doesn't exist? Perhaps it is time for a "first cut" at some potential misunderstandings.

D. POTENTIAL MISUNDERSTANDINGS: A FIRST CUT

There are many ways to misunderstand the systematic distortion hypothesis. Some of these misunderstandings will now be addressed.

1. Raters Are Not "Blind"

It would be a mistake to conclude that raters are typically ignorant of the persons rated, that their responses are fabricated, or that they are typically insensitive to the intensity of a signal. Indeed, most raters know a lot about the person rated and they can bring to mind some combination of recalled incidents and prior attributions. It is not what raters know that gets them into trouble, but rather the inferences they are prepared to draw starting from retrieved knowledge that is probably unrepresentative, on the one hand, and ending with generalizations that follow pathways of conceptual association rather than probabilistic association, on the other.

The point of the systematic distortion hypothesis is not that particular rater judgments are characteristically inaccurate. Indeed, in the Newcomb (1929) data, ratings for single variables correlated in the $.40$ -. $.50$ range with measures derived from daily records. The real point is that (a) raters are far from perfect and, when raters do make errors, their errors are systematically biased, not random; (b) the typical memory-based personality rating situation requires the rater to abstract and summarize a mass of observations from hours, 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; and (c) 20 or 30% error variance on particular items is probably enough to support moderate size, illusory interitem correlations. The systematic distortion hypothesis does not imply that raters are insensitive to a signal or only see what they expect to see!

2. "How People Classify," Not "Who Gets Classified How"

To understand the systematic distortion hypothesis one must honor the distinction between "classification" and "identification." Following Sokal (1974), classification can be defined as "the ordering or arrangements of objects into groups or sets on the basis of their relationships" (hence, the concern for correlational structure); identification can be defined as "the allocation or assignment of additional unidentified objects to the correct class, once such classes have been established by prior classification." The systematic distortion hypothesis is concerned with "how people classify" (classifications), not "who gets classified how" (identifications) (D'Andrade, 1965).

Classification and identification are different cognitive activities. Scientific classifications are designed to summarize the relationships among variables ("things that are hot" are "things that hurt," people who are "friendly" are "not aggressive") (Gilmour, 1937) and thus help us make valid inductive generalizations. But, as Hakeel (1971) notes, sharing a structure of intervariable inferences among personality variables (classification) does not necessarily standardize agreement on the personality attributes that apply to rates (identification). "Two raters might have identical patterns of trait intercorrelations and show no agree-

ment (on whether such-and-such rating category applies to such-and-such race) ($r = .00$) or even perfect disagreement ($r = -1.00$). "The systematic distortion hypothesis is concerned with the scientific status (internal homogeneity and external validity) of the categories into which people are classified (e.g., introversion: hesitant to speak before a group, does not like parties, prefers to work alone) not with the question whether Aunt Sally is an "introvert."

III. The Systematic Distortion Hypothesis and Personality Structure

The systematic distortion hypothesis is a critique of the validity of personality classifications induced from memory-based personality ratings, implicit personality theory, or conceptual association judgments. The hypothesis implies that if one were to eliminate error and systematic bias from memory-based personality data, one would not discover global traits, broad syndromes, or general factors, but rather a complex of context-dependent truths, person by situation by response mode statistical interaction effects, and unstable intercorrelations among events, or alternatively said, that which is accurate in personality ratings would not support a global trait approach to individual differences in conduct.

A. THE GLOBAL TRAIT MODEL

A global trait theory of personality is an explanatory theory of a particular kind. It sets as its goal the reduction of the complexity and diversity of individual differences over time and context to a limited set of underlying forces in terms of which individuals may be said to differ regularly, a limited set of "more or less stable internal factors that make one person's behavior consistent from one time to another, and different from the behavior other people would manifest in comparable situations" (Child, 1968, p. 83).

A global trait theory is successful if its categories can be applied to make the diversity of individual differences appear consistent across situations, response modes, and time. Global trait theorists commit themselves to each of the following claims.

1. People differ in their behavior in comparable situations. (*Note:* the comparability of the situations in which differences in behavior are observed is crucial. Global trait theorists are not interested in why some people talk more in debating clubs than other people talk in libraries.)
2. Within any given situation there are behaviors which distinguish people from one another, and there are behaviors which do not distinguish people from one another. That is to say, investigators can separate behaviors typical of the situation, things anyone would do "under the circumstances," from the be-

haviors with respect to which individual differences are observed in that situation. The fact that (e.g.) some students take notes during a lecture while others do not is grist for the mill of the global trait theorist. The fact that all students sit down in their seats during the lecture is *not* grist for the mill of the global trait theorist.

3. For any given person there are two sets of behaviors. One set consists of those individual difference behaviors that are characteristic of a particular person in the situations in which his or her behavior has been observed. The other set consists of those individual difference behaviors observed in the situations in which his or her behavior has been observed that are not characteristic of the person's behavior (i.e., behavior characteristic of other people's behavior in those situations).

4. The behaviors in both sets can be shown to be reducible to a relatively small set of theoretical categories which reveal the broad underlying generality of behavioral differences. For example, think of all the diverse behaviors (walks alone, tends to depression, easily hurt, worries over possible misfortune, reads a lot, suspicious of the motives of others, attends to details, daydreams, avoids talking before a group) that are theoretically equated and unified by the single underlying category, "introversion."

The first three propositions listed are endorsed by all individual difference theorists, and remain uncontroversial. The fourth proposition, however, is distinctive of global trait theorists, and that is where the arguments begin (e.g., Mischel, 1968, 1973; Raush, Dittmann, & Taylor, 1959; Mow, 1969; Shweder, 1973, 1979a; Nisbett, 1980).

Global trait theorists seek to reduce the diversity of individual differences in behavior to a limited number of underlying factors (proposition 4). This is done empirically by detecting redundancies in the correlational structure of individual difference variables across response modes (the extroverted person is more likely to talk, make friends, and lend his possessions), and situations (the extroverted person is more likely to make friends at a party and in the dormitory). For the most part, the correlational redundancies among personality variables required to support global trait theory have been retrieved from response patterns on memory-based instruments. Child psychologists concerned with child-rearing practices speak of mothers who are "permissive" and point to a redundant cluster of correlated scales on a socialization questionnaire (e.g., Sears *et al.*, 1957). Clinical psychologists concerned with psychopathology speak of persons who are "ego-resilient" and point to a correlated pattern of scales on the MMPI (Black, 1965). Let us depart for a moment from this tradition of constructing personality taxonomies out of test responses, especially since what we are discussing is systematic bias in memory-based procedures.

Presumably, personality researchers became interested in rating and inventory behavior not because they were interested in this type of behavior per se, but