FACT AND ARTIFACT IN TRAIT PERCEPTION: THE SYSTEMATIC DISTORTION HYPOTHESIS

Richard A. Shweder
COMMITTEE ON HUMAN DEVELOPMENT
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS

I. Introduction: The Systematic Distortion Hypothesis


*This article is a revision and synthesis of sections of two unpublished manuscripts entitled “Fact and Artifact in Personality Assessment: The Influence of Conceptual Schemata on Individual Difference Judgments,” and “Attributional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 Forsaken 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 relational structure from a 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," "agrees" 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 exist 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; Eibsen & Allen, 1977, 1979; Cooper, 1981; Geis & Rosenberg, 1981; Mirch, n.d.; Nisbett, 1980; Borman, 1982) and criticized in others (Block, 1977; Block, Weiss & Thorne, 1979; Epstein, 1979; Jackson, Chan & Stricker, 1979; Luminell, 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

---

**Fig. 1.** The systematic distortion hypothesis. Degrees of correspondence ($r_d$) 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 similarities 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., "disagree" and "criticize") do not parallel the correlational structure of actual behavior (e.g., "disagree" and "criticize" 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 cooccurrence 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 differentiatation 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.
It is noteworthy that "impressive" personality taxonomies (e.g., LaForge & Succzboth, 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 interpersonal 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 (MDSACR) (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 (MDSACR) (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 of .80, .86, and .70 between the three rated behavior 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:

### TABLE II

**Bales’ Three Dimensions of Personality and Interpersonal Behavior**

<table>
<thead>
<tr>
<th>Item number</th>
<th>Power</th>
<th>Likeability</th>
<th>Task-orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conceptual dimension</td>
<td>Rating dimension</td>
<td>Conceptual dimension</td>
</tr>
<tr>
<td>1</td>
<td>- .30</td>
<td>.01</td>
<td>- .30</td>
</tr>
<tr>
<td>2</td>
<td>.18</td>
<td>.34</td>
<td>.74</td>
</tr>
<tr>
<td>3</td>
<td>.71</td>
<td>.77</td>
<td>.82</td>
</tr>
<tr>
<td>4</td>
<td>.43</td>
<td>.65</td>
<td>- .00</td>
</tr>
<tr>
<td>5</td>
<td>-1.13</td>
<td>- .99</td>
<td>- .06</td>
</tr>
<tr>
<td>6</td>
<td>- .13</td>
<td>- .15</td>
<td>.89</td>
</tr>
<tr>
<td>7</td>
<td>.62</td>
<td>.66</td>
<td>.41</td>
</tr>
<tr>
<td>8</td>
<td>.43</td>
<td>- .34</td>
<td>- .68</td>
</tr>
<tr>
<td>9</td>
<td>- .09</td>
<td>.03</td>
<td>- .44</td>
</tr>
<tr>
<td>10</td>
<td>.65</td>
<td>- .01</td>
<td>.89</td>
</tr>
<tr>
<td>11</td>
<td>- .23</td>
<td>.16</td>
<td>- .53</td>
</tr>
<tr>
<td>12</td>
<td>.00</td>
<td>- .02</td>
<td>.21</td>
</tr>
<tr>
<td>13</td>
<td>.65</td>
<td>.64</td>
<td>.03</td>
</tr>
<tr>
<td>14</td>
<td>.38</td>
<td>.10</td>
<td>1.04</td>
</tr>
<tr>
<td>15</td>
<td>-1.22</td>
<td>-1.14</td>
<td>- .09</td>
</tr>
<tr>
<td>16</td>
<td>- .71</td>
<td>- .72</td>
<td>- .23</td>
</tr>
<tr>
<td>17</td>
<td>.45</td>
<td>.51</td>
<td>- .55</td>
</tr>
<tr>
<td>18</td>
<td>- .29</td>
<td>- .41</td>
<td>.61</td>
</tr>
<tr>
<td>19</td>
<td>-1.12</td>
<td>-1.28</td>
<td>.26</td>
</tr>
<tr>
<td>20</td>
<td>- .70</td>
<td>- .46</td>
<td>- .09</td>
</tr>
<tr>
<td>21</td>
<td>.79</td>
<td>.75</td>
<td>.06</td>
</tr>
<tr>
<td>22</td>
<td>.73</td>
<td>.79</td>
<td>.85</td>
</tr>
<tr>
<td>23</td>
<td>.47</td>
<td>- .47</td>
<td>.45</td>
</tr>
<tr>
<td>24</td>
<td>- .05</td>
<td>.34</td>
<td>- .83</td>
</tr>
<tr>
<td>25</td>
<td>.34</td>
<td>.40</td>
<td>1.11</td>
</tr>
<tr>
<td>26</td>
<td>.76</td>
<td>.90</td>
<td>- .78</td>
</tr>
</tbody>
</table>

r_{p} of conceptual dimension and rating dimension: 86, 88, 70

---


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 [accurate 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 noncorrelational 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 (1977) 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 dis-
similar in meaning either because the judge cannot imagine applying both terms to
the same act, or because the two terms evoke opposed connotations (e.g., 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" (e.g., 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
decided 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 & Stovin, 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, 1977) 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 association methods (e.g., Szalay & Deese, 1978) provide a second
way to choose between systematic distortion and accurate reflection interpreta-
ations of similarity of meaning judgments. The systematic distortion hypothesis
predicts that direct pairwise similarity of meaning judgments should be reproduc-
able 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 judg-
ments 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 descrip-
tions 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 judg-
ments. The two intermatrix correlations were .61 and .72. D'Andrade (1974) and
Shweder (1977e) have argued that conceptual association matrices reflect a va-
riety 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 Steffire, Reich, and
McClaren-Steffire (1971) and D'Andrade, Quinn, Nerlove, and Rommey (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 interim similarity structure of conceptual association matrices with the interim correlational structure of actual behavior matrices. To date comparisons with seven data sets have been performed (D'Andrade, 1974; Sweder, 1975, 1977a; Sweder & 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), Sweder (1975, 1977a), and Sweder 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 Sweder 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 (Sweder & 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 type of behavior as soon after its occurrence as possible. For the Sweder 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 Sweder & 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 (Sweder, 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 subjective 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; Sweder, 1975, 1977a; Sweder & 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 rates is unrelated, or opposed, to what associates with what in
the preexisting conceptual association scheme of raters. Will the interim correlational structure of the (subsequently rated) rated behavior matrix replicate the interim proximity structure of the preexisting conceptual association matrix, as predicted by the systematic distortion hypothesis, or will it correspond to the interim 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 interim proximity structure of preexisting conceptual associations was in radical conflict with the interim correlational structure of observed individual differences. Under these manipulated circumstances, what happened to the interim correlational structure of memory-based ratings? Memory-based ratings of what correlated with what replicated the interim proximity structure of preexisting conceptual associations ($r = .84$) and were insensitive to the interim 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 interim conceptual associations and interim 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 associations 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 (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, system bias, occur; and (c) 20 or 30% error variance on particular items is probably enough to support moderate size, illusory interim correlations. The systematic distortion hypothesis does not imply that raters are insensitive to a signal or only see what they expect to see.


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 Hake (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-
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 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 beh-

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; Meeus, 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 (Block, 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
B. THE FAILURE OF THE GLOBAL TRAIT MODEL

The systematic distortion hypothesis challenges us to construct a model of personality structure with data from on-line records of conduct instead of memory-based assessment procedures. When one shifts methods from memory-based measures to behavioral observational techniques, one discovers that the world of individual differences is not organized in the way envisioned by the global trait theorists. Individual differences exist (propositions 1-3), of course, but they do not seem to generalize widely across similar response modes or situations.

There are relatively few homogeneous multitem clusters (e.g., dependent: seeks help, seeks attention, seeks physical proximity) that can be induced from behavior observational data on individual differences. Most personality researchers who collect data using on-line, immediate scoring techniques tend to discover complex person by response mode by situation statistical interaction effects. The observed set of individual difference behaviors characteristic of most persons (see Section III.A, proposition 3) looks (e.g.) as follows: speaks well in public, slow to make friends, prefers to work alone, sociable at parties. Such sets of behavior are not easily reducible to a small set of underlying global traits. Stated alternatively, when individual differences are observed across theoretically similar response modes (e.g., seeks attention, seeks help), and across naturally occurring situations (e.g., at home and at school), hypothesized global traits (e.g., "dependency," "assertiveness," "risk-taker") typically account for no more than 10% of the diversity of individual differences. The financial "risk-taker" is not the social "risk-taker," and the child who seeks help is not typically the one who seeks physical nearness (see, e.g., Newcomb, 1929; Raush et al., 1959; Raush, Farberman, & Llewellyn, 1960; Sears, 1963; Longabaugh, 1966; Mischel, 1968; Moos, 1969; Slovic, 1972; Shweder, 1973, 1979a; Whiting & Whiting, 1975, p. 163; Cronbach, 1975; Yarrow & Waxler, 1976; Fiske, 1978).

Our ability to predict individual differences is substantially augmented if response mode and context are taken into account. One implication of this fact is that scientific personality taxonomies cannot be directly borrowed from the global abstract trait lexicon of implicit personality theory (e.g., "dependent," "friendly," "honest," "responsible"). Since high-order statistical interaction effects (time, place, response mode, dyad, etc.) prevail, scientific personality classifications must concretely specify the response mode and the contextual conditions under which correlational patterns can be expected to be found. Indeed, from a behavior observational point of view, one should probably not speak of "children who are dependent," but rather of "children who seek attention," on the one hand, and "children who seek physical nearness," on the other. And, if one were to find subsequently that "children who seek attention from their mothers in the playground" are not more likely than other children to "seek attention from their mothers at home," one would again want to revise one's taxonomic categories by writing in more context (reference to time, place, personnel, etc.).

The systematic distortion hypothesis challenges us to confirm global trait personality taxonomies with systematically collected, reliable, on-line behavior observational evidence. When Newcomb tried this in 1929, he discovered that individual difference variables that should have correlated according to the global trait theory were no more significantly intercorrelated in actual behavior than a random selection of variables. Shweder (1973) discovered a similar thing in an examination of reported behavior observational intercorrelations among various indicators of "nurture" and "egoism." Global trait categories like "nurture" generate a theoretical expectation that (e.g.) children who "offer help" more than others are more likely than others to "make responsible suggestions" and less likely than others to "seek help." An examination of on-line data collected by L. Minturn (reported in Longabaugh, 1966) found little support for these theoretical expectations.

Global trait predictions generally have not fared well in the confrontation with behavior observational evidence. Therefore, one goal of global trait theorists probably cannot be realized, viz., to derive a parsimonious taxonomy that will enable them to predict validly individual differences in one context and response mode from knowledge of individual differences in another context or response mode. As Raush et al. (1960) remark, it is the "idiocentric" or "interactive" effects, "the particular meaning that a particular situation has for a particular person that is the major determinant of behavior." Memory-based ratings incline to the nomothetic; behavior observational evidence inclines to the idiographic.

C. POTENTIAL MISUNDERSTANDINGS: A SECOND CUT

There are many ways to misunderstand what has just been said about personality structure and the systematic distortion hypothesis. Some of these potential misunderstandings will now be addressed.
1. Personality Does Exist

The systematic distortion hypothesis does not deny the existence of trait dispositions. It does deny the existence of global traits, as retrieved from memory-based assessment procedures. The issue is not whether traits or individual differences exist, but rather how widely they generalize. We may well live in a world where an internally homogeneous, stable trait 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 that 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.

2. Situationalism Is No Alternative to Personality Theory

The idea that global traits do not exist and that individual differences are narrowly context-dependent does not imply that situation or context is all there is to behavior. That would be tantamount to denying the existence of individual differences within situations, which would be ridiculous. Moreover, it is probably the case that, just as personality dispositions are context-dependent, situational influences are person-dependent (e.g., Raush et al., 1959; Moos, 1969). The way situations differ in their ability to elicit certain kinds of behaviors (e.g., cooperation, aggressiveness) typically depends on which person is involved. Statistical interaction effects are multiple necessary conditions (Kelley, 1972). Broad “main effects” are relatively rare for situations as well as persons.

3. There’s More to Personality than Global Traits

The systematic distortion hypothesis does not presuppose that the only legitimate goal for personality researchers is to predict individual differences in one context from knowledge of individual differences in other contexts. The aims of the nomothetic taxonomist are not coincidental with the goals of personality research. Indeed, idiographic, clinical insight often involves an appreciation of unique context-person-response mode interactions and the acquired (and possibly unconscious) meanings of specific stimuli for specific persons (see Shweder, 1980b). There is more to personality than global traits.

4. Factor Analysis Is a Fine Thing

It would be a misunderstanding of the systematic distortion hypothesis to conclude that factor analysis, multidimensional scaling, and other statistical procedures for arriving at global multitem categories are of limited usefulness in the study of human conduct. The implications of the systematic distortion hypothesis are restricted to personality classification per se, that is, the way individuals differ from each other across a variety of response modes and contexts as revealed by immediate scorings of conduct. It is quite possible there are global factors or dimensions that can be inferred from behavioral records when one switches from studying individual differences to studying the way behaviors correlate across units such as dyads, behavioral episodes, and groups (see Shweder, 1972, 1973).

5. There Are No “Immaculate Perceptions”

The systematic distortion hypothesis is a critique of the validity of item-item correlational structures retrieved from memory-based personality ratings. The validity of rated behavior matrices is called into question by comparing their correlational structure to that of immediate, on-line scorings of conduct. The appeal to behavior observational evidence does not imply that theory, interpretation, and judgment should be banished from personality research, nor does it suggest that immediate, on-line scorings of conduct are “immaculate perceptions.” The two critical assumptions made by advocates of the systematic distortion hypothesis are (a) what on-line scorers mean by (e.g.) “criticizes,” or any other coding category is not different from what memory-based raters mean by the same category; and (b) reliable on-line scorings are a closer approximation of the correlational structure of actual behavior than are memory-based ratings.

On-line scorings are not interpretation-free; behavior is not self-describing. There are many ways to “code” behavior—each coding scheme transforms a behavioral event into a theory-laden significance. Is behavior to be described in terms of muscle twitches or actor intentions? If intentions are to be inferred, how much context should the coder appeal to in inferring the intention? Should the coder appeal to the outcome of the act, its effects on others, the actor’s stated goals, events from earlier that day, earlier that year, or perhaps events prior to the actor’s own recall (e.g., events from the first few years of life that unconsciously motivate the coded event)? Different coding schemes “translate” the stream of behavior in different ways. The systematic distortion hypothesis assumes only that whatever coding scheme is used, on-line scorers speak the same “language” as memory-based raters.

6. Some People Do Have Global Traits

The systematic distortion hypothesis implies that, given most global trait categories or factors (e.g., dependent vs self-reliant, egotistical vs altruistic), most people will mix together in their behavior items that should not go together from the point of view of the trait category. This implication does not deny that some people display global generality. Indeed, it may also be the case that for any particular person one or more aspects of his or her personality are characterized by global generality. The main difficulty emphasized by the systematic distortion hypothesis is that, from a nomothetic point of view, one can predict too few of the people too much of the time (pace Bem, 1974).
7. Aggregation Is Not the Answer

The idea of a global trait (e.g., extroversion) is that individual differences generalize widely across similar response modes (talkative, energetic, organizes activities) and across comparable situations for equivalent response modes (talks to his wife at dinner, talks to his or her colleagues at faculty meetings). The systematic distortion hypothesis states that memory-based evidence of internal homogeneity among trait indicators cannot be trusted, and that behavior observational evidence will not reveal the pattern of interitem correlations necessary to support the global trait concept.

Do global traits emerge from aggregated behavior observational evidence? Data are aggregated or pooled primarily to compensate for a "noisy" channel (measurement error due to faulty instruments), or to amplify a weak signal. Most of the observational evidence referred to in this article is derived from data aggregated over several occasions (see Nisbett, 1980, for a review). Newcomb's (1929) data, for example, are derived from 24 days of observation with odd- vs even-day reliabilities of .78. For obvious reasons, however, the data are aggregated across repeated occasions, but not across similar response modes or equivalent response modes for different situations. According to the systematic distortion hypothesis, similar response modes and equivalent response modes for different situations are not sufficiently intercorrelated to support the idea that people have global traits consisting of covarying behaviors. To aggregate data across similar response modes and across different situations would be tantamount to canceling out by methodological fiat precisely those statistical interactions of person-situation-response mode that are embarrassing to the global trait concept (see Mischel & Peake, 1981).

IV. The Systematic Distortion Hypothesis and Implicit Personality Theory

Within personality psychology there are two widespread beliefs about implicit personality theory. The first belief is that ordinary language dictionary entries for personality and interpersonal behavior (e.g., industrious, honest, cooperative, dependent) are lexical labels for quasi-scientific categories. The descriptors in our everyday lexicon are often appropriated into academic discourse as though they were inductive summary formulas about personality dispositions which had utility in predicting future events and minimizing surprise (Brown, 1965, p. 612). The trait lexicon as a whole is often viewed as a "coding system" for "packaging" information about the correlational structure of individual difference variables (Cattell, 1943; Bruner, Shapiro, & Tagiuri, 1958; Passini & Norman, 1966; Jackson et al., 1979). Utilizing this ordinary language lexicon as a coding system, ordinary folk are thought to develop more or less accurate implicit personality theories about what correlates with what across individual differences in conduct (e.g., people who are "self-reliant" or "responsible").

The second prevalent belief in personality psychology is that if our ordinary language coding scheme yields inaccurate knowledge about what goes with what in personality, then ordinary language terms (e.g., self-reliant, aggressive, friendly, responsible, dependent) are either linguistic artifacts of no genuine importance, or should disappear from the language. I am doubtful of both beliefs.

One implication of the systematic distortion hypothesis is that ordinary language trait terms for describing personality and interpersonal behavior are too undifferentiated, too global to be scientifically useful. Individual differences are not stabilized and packaged at the abstract level of "dependent," "honest," and "hostile," but rather at concrete levels such as "tends to seek help when adults are around," "cheats at poker," or "gets irritated if evidence for global traits cannot be retrieved from behavior observational evidence." A scientifically useful coding scheme for personality differences must be sufficiently differentiated to encode individual differences. This may be done by means of response mode descriptors (e.g., "seeks help") and situational descriptors (e.g., "from male adults").

The inferences one can draw from our ordinary trait lexicon are not useful for predicting future events and minimizing surprise. It does not follow, however, from this implication of the systematic distortion hypothesis that our ordinary trait lexicon has no important function in human affairs, or is a linguistic artifact on the wave. It does not follow because there is no reason to postulate that ordinary language terms have evolved to serve only the goals of inductive science, no reason to assume that the activities in which trait descriptors play a part in everyday life are only scientific activities, and no reason to reduce the uses of language to the aims of science. Our trait lexicon may well serve important functions other than summarizing and encoding correlational structure.

A. The Uses of a Trait Lexicon: Scientific vs Nonscientific Functions

When are ordinary language trait terms actually used? At what types of occasions is personality trait talk to be found? For the moment these questions cannot be answered with any confidence; we know little about the way our trait lexicon is actually used in everyday life. Thus, I can only speculate and look forward to future research.

What little do we know about the scientific and nonscientific uses of our ordinary trait lexicon? A priori, one can imagine nonscientific functions. Most trait terms can be used as adjectives of appraisal communicating either approval ("thrift") or critical evaluation ("stingy"). Trait terms can be used to inspire
conduct ("be courageous"), to proscribe conduct ("stop being so dependent"), or to influence the way others will feel about or react to a person ("he's dishonest"). Given the diverse functions of language—the rhetorical, the persuasive, the regulatory, the symbolic—there seems to be no a priori reason to assume that trait terms evolved to serve scientific functions exclusively.

There are other reasons to doubt that our ordinary trait lexicon evolved as a scientific coding scheme. Ordinary folk do not need global trait terms to describe one another's personalities and do not always use global terms. In one cross-cultural study of free descriptions of personality, for example, Shweder and Bourne (1981) discovered that 46% of American descriptors and 29% of Oriya (Indian) descriptors are unqualified global trait attributions. Eighty percent of Oriya descriptors (and 54% of American descriptors) referred to either response mode ("she shouts profanities") or context ("she brings cakes for my children on festival days"), and this despite the fact that the Oriya language is rich in global trait descriptors and Oriya informants are capable of generating global trait terms. It is by no means obvious that interpersonal description in everyday life is typically executed at the abstract level of our trait lexicon.

Moreover, ordinary folk do not typically explain one another's conduct by means of our personality trait lexicon. For example, in one quasi-naturalistic study, Lewis (1978) asked informants to explain why someone in their immediate environment was doing whatever he or she was doing. Lewis discovered that less than 5% of everyday explanations of conduct (why he is out jogging today?) is in global trait form (e.g., "he's compulsive"). He discovered that everyday explanations for actions are typically goal-oriented (e.g., "he's studying because he has an exam tomorrow") and everyday explanations for emotions are typically situationalist—emotions are thought to be externally elicited, not voluntarily willed (e.g., "she's feeling nervous because the highway is covered with ice"). Global personality trait explanations seem to be relatively rare in everyday life.

When is our global trait lexicon utilized? More research has to be done before this question can be answered with any confidence. Speculating, I would hold out "personnel selection" as the paradigmatic "language game" within which global trait terms are bandied about. By "personnel selection" I mean any occasion that requires justifying the allocation of individuals to positions in social groups and social networks.

Members of functioning social groups must answer certain perennial questions about the recruitment of members, the selection of leaders, and the formation of bonds of intimacy, cooperation, and sharing. Who's in and who's out? Who's up and who's down? Who's close and who's distant? (Schatz, 1967; Bales, 1968, 1970). All of us, as members of collectivities, routinely make decisions about who should be accepted, promoted, retained, confided in, allied with, delegated responsibility, etc. These are the personnel selection decisions of everyday life.

My speculation suggests that around such decisions global personality trait talk plays a major part (in gossip, recommendations, formal decision making, etc.). This should be true in all societies, but especially so in a voluntaristic, individualistic culture where entrenched values make it reprehensible to accept, promote, embrace, or befriend on the basis of family connections, birthright, caste, color, etc. Thus, it is not surprising that underlying group structure, role perception, and personality attribution is a universal semantic space for conveying messages about hierarchy-subordination and solidarity-conflict (Bales, 1970; Shweder, 1972; Kirk & Burton, 1977; White, 1980), a common semantic space that makes it possible to "map" personality trait labels (assertive, self-esteem, self-reliant) onto social positions (e.g., "leader").

Why should personnel selection be carried on with terms from our global trait lexicon? One conceivable answer is that global trait attributions predict role performance. Perhaps information about peer or interviewer impressions coded in global trait terms (self-reliant, responsible, honest) reduces uncertainty about how people will conduct themselves as leader, friend, ally, subordinate, etc. There is not much convincing evidence to encourage us in this view.

Global trait attributions are notoriously unreliable. Indeed, merely to attain respectable split-half interobserver reliabilities from lay observers using our global trait lexicon as a coding scheme one must pool the attributions of approximately eight observers (Smith, 1967), a procedure rarely followed in everyday life. Pairwise interobserver reliabilities seem to range from .15 to .35 (Hake, 1971; Fiske, 1978). Indeed, Bourne (1977) discovered that pairs of observers of the same person agreed only about 20% of the time in their attributive selection of global trait labels. Observers of two different people agreed 16% of the time. The difference in amount of agreement, two observing one and two observing two, was not significant.

Moreover, role performance is probably quite sensitive to "treatment effects" (Einhorn & Hogarth, 1978; Einhorn, 1980). Those who are accepted, selected as leaders, made intimates, etc., are given opportunities and treated in ways withheld from those who are rejected, kept at a distance, etc. It is much easier to be cooperative when someone holds out a hand. It remains to be seen whether the allocation of people to social positions, solely on the basis of the global trait impressions of three or four peers, yields significantly more effective leaders, cooperative allies, loyal employees, etc., than a selection procedure among qualified recruits based on a random number table or astrological chart. My suspicion is that the results would not justify a heavy investment in the attributive use of our global trait lexicon.

Thus, I am led to consider a second conceivable answer to the question: Why should personnel selection be carried on with terms from our global trait lexicon? Perhaps personality trait talk serves a symbolic function, not a predictive one. Perhaps personnel selection is a gloss for an expressive ritual in which we say
something with words and action about our cultural values, concepts of the person, and views of man in society (for a discussion of expressive rituals, see Shweder, 1981). If groups are to function, decisions must be made about how to distribute individuals to positions in social groups and networks. But, if interobserver reliability is low, interitem correlations among indicators of the “same” global trait are weak, and treatment effects are substantial, it is conceivable that, apart from gross screening, from a predictive point of view, almost any selection device will do. Note, however, that not any selection device will do from a symbolic point of view.

The principles we use to allocate individuals to social roles express something about our values, goals, and view of the “person.” A culture like our own, deeply committed to the axiom or premise that society derives its authority from the consent of autonomous and unique individuals, takes an interest in the search for an underlying, stable core character, and has reason to subscribe to the fiction that personality fits the role (see Selby, 1974, 1975, for an alternative, Zapatos view). Neither a random number table nor a genealogical tree are the right kind of symbol for the image of man enshrined in our culture. Indeed, some of the best predictors of role performance, race, religion, and sex are ignored because they imply social classifications symbolically offensive to our axiomatic notions of justice, personal achievement, and individualism.

It should be reemphasized that much of what has just been said is speculation designed to encourage future research. One central point, however, should not be lost in the midst of this speculation: Ordinary language global trait terms are not necessarily labels for scientific categories. Our global trait lexicon may serve functions other than packaging information about the correlational structure of individual differences in conduct.

It is hoped that the implication of this point is also clear: The persistence of global trait categories in ordinary language may be unrelated to issues of truth or falsity. One should not expect trait terms to go away simply because they fail 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.

B. POTENTIAL MISUNDERSTANDINGS: A THIRD CUT

There are a number of ways to misunderstand what has been said about lexical categories, the scientific and nonscientific functions of language, and implicit personality theory. Some of these potential misunderstandings will now be addressed.

1. Global Trait Terms Are Not Linguistic Artifacts

The systematic distortion hypothesis does not imply that our ordinary language lexicon of global trait terms (e.g., responsible, dependent) comprises merely labels for deficient scientific categories or linguistic artifacts of no importance. It is certainly true that in the hands of some personality trait theorists our ordinary language trait lexicon has been appropriated for scientific use, for which it is probably deficient. It is also true that ordinary folk sometimes use global trait terms for interpersonal description, although global trait terms are rarely used to explain behavior. However, language serves many functions: descriptive, prescriptive, rhetorical, persuasive, regulatory. Although global trait categories are poor scientific categories this does not mean that our trait lexicon has no important place in our culture.

2. Language Does Not Correspond to Reality

The systematic distortion hypothesis implies that the various referents (e.g., seeks help, seeks attention, seeks proximity) of a global trait term (e.g., dependent) do not typically hang together, empirically cohere, or intercorrelate across personalities; likeness is not paralleled by cooccurrence likelihood. Some critics of the systematic distortion hypothesis treat this implication as a denial of a premise they take to be self-evident, viz., that words are related to, or correspond to, things in reality. The language-reality correspondence assumption is probably widespread. Before debating its implications for the systematic distortion hypothesis it may be useful to clarify what it means. Upon examination, the notion of language-reality correspondence seems to me to be a premise without promise.

The idea that words are related (or correspond) to things in reality is an empty notion, until one specifies the particular way words are supposed to be related to things in reality. It is necessary to be specific because there are obviously many ways words are not related to things in the world many ways in which the idea of language-reality correspondence is plainly false. No one, for example, would ask how the world is spelled (see Goodman, 1960), nor would anyone suggest that the sequential order of words in a sentence is paralleled by some sequential order of the things mentioned by the words in a sentence, nor would anyone suggest that big words (e.g., “microorganism”) are names for big things. So, in what way are words related to things in reality?

It is, of course, quite undeniable that words are related to things in the world in at least one sense: Words are things in the world. But that is a trivial point.

Words are also related to things in the world in a second sense: Words refer to things in the world, although some of the things in the world they refer to are other words, symbols, and representations. If all words have reference, then some of the things they refer to must be other words, symbols, and representations. This qualification is necessary if one is to explain the difference in meaning of such terms as “poo-poo,” “feces,” and “shit,” all of which refer to the same nontangible thing in the world, or the difference in meaning of terms such as “elf,” “centaur,” and “unicorn,” all of which refer to nothing except perhaps
things in the world like descriptions, stories, or drawings of "elves," "centaurs," and "unicorns" (see Goodman, 1949, 1968).

Not much follows, however, from the fact that words have reference. Certainly it does not follow that the things referred to by words are like the words themselves or correspond to them. A knee, for example, is part of a body. Does it follow that the word "knee" is part of the word "body"?

But, perhaps I have missed some subtlety in the apparently naive premise that language corresponds to reality. Fortunately, the claims of the systematic distortion hypothesis are less subtle: (a) There are many true things one can say about pairs of behavioral descriptors (e.g., the behaviors referred to by the descriptors have overlapping reference, the behaviors follow one another in a script, the behaviors are both parts of an image of a behavioral type, the behaviors are pleasant, sexy, and powerful); (b) most of the true things one can say about pairs of behaviors reveal little about the joint probability of occurrence of the behaviors across personalities; and (c) judges on memory-based assessment procedures confuse one set of truths (likeness, sequential dependency, part-part, common effect, etc.) with a second set (i.e., cooccurrence likelihood).

Fortunately, the distinction between likeness and likelihood is honored by most of us, even when, in our less vigilant moments, we slip into the careless assumption that language corresponds to the world. For if language and conceptual associations did imitate reality, we would not have to observe behavior at all. Face validity would guarantee internal and external validity. That would be very strange, indeed. Unfortunately, that seems to be what has happened in the personality psychology literature. Sets of items that have face validity conceptually associate in the minds of raters. Raters report correlational structures that replicate their preexisting associative networks, thereby producing illusory evidence of internal and external validity. Propositions about likeness are confused with propositions about cooccurrence likelihood. Ideas become reality.

3. Ordinary Folk Do Have Valid Social Knowledge

The systematic distortion hypothesis raises doubts about our knowledge of correlational structure and suggests that our correlational beliefs are not always in rapport with the contingent structure of our environment. The hypothesis does not imply, however, that we lack valid social knowledge or that valid correlational beliefs about individual differences are necessary for successful adaptation in a social environment.

The concept of correlation is a formal operational one (Inhelder & Piaget, 1958); it calls for a comparison of two conditional probabilities. Recent research suggests that normal, intelligent adults are disinclined to formal operational thinking (Wason & Johnson-Laird, 1972; Tversky & Kahneman, 1974) and have difficulty processing correlation-relevant information (Smedslund, 1963; Ward & Jenkins, 1965; Shweder, 1977c; Shweder & D'Andrade, 1980). It seems to me that one implication of this research is that many adaptive processes do not require formal operational thinking, and may only require limited intellectual skills. In everyday social life, one can get quite far with the ability to imitate, to notice things that are contiguous in time or space, to recognize things one has seen before, and to estimate relative frequencies. Valid correlational thinking may not be a requisite for social adaptation.

Ordinary folk do possess valid social knowledge and much of the social knowledge they possess requires relatively low-level, mundane inference abilities. Some of this knowledge is episodic and script-like (Schank & Abelson, 1977; Forgas, 1979); we know the social routines and sequential steps for "going to the restaurant," "greeting a friend," "interviewing a prospective employee," or we know how to find out about the routine (e.g., "buying a house"). Some of the knowledge is repetitive and context-specific; we know how so-and-so behaved last time under such-and-such circumstances (past behavior predicts future behavior), and we know the kinds of things so-and-so wants and how to manipulate them to get him or her to do what we want (context-specific knowledge of "reinforcers"). There is no reason to assume that our potential for adaptive behavior must be derived from some supposed formal operational ability. The context-specific intellectual demands of everyday social life may not be very great.

V. Conclusion: From Likeness to Likelihood

The systematic distortion hypothesis casts doubt on the belief that people have global personality traits consisting of multitem covarying behaviors. The hypothesis implies that individual differences are not organized in the way envisioned by global trait theorists. According to the systematic distortion hypothesis, individual differences do not generalize very widely, that is, there are few empirically homogenous multititem categories (e.g., "aloof," "dependent," "extroverted") that can be induced from the interitem correlational structure of behavior observational data on individual differences.

The systematic distortion hypothesis states that judges on memory-based personality assessment instruments (inventoryes, rating forms, questionnaires) either infer correlational structure from a general model of conceptual association, or find conceptually associated memory items easier to retrieve. Conceptual proximity is utilized as a guide for estimating cooccurrence probability, thereby replicating a preexisting associative network that has little to do with what correlates with what across personalities. Likeness is substituted for likelihood.

Thus, according to the systematic distortion hypothesis, correlational clusters (i.e., global traits) derived from memory-based procedures cannot be treated as accurate reflections of the empirical correlational structures of behavior.
The systematic distortion hypothesis challenges us to anchor our claims about the organization of individual differences in the firmer ground of reliable, on-line behavioral observational evidence. There would, of course, be little point in utilizing data-gathering techniques which require that we engage in the arduous and time-consuming enterprise of observing, immediately recording, and coding a heavy dose of behavioral occurrences, if we could rely on raters to abstract and summarize the essentials from masses of behavioral events. Unfortunately, human memory and inference are subject to many influences, one of which is the structure of conceptual associations among the descriptors we use to assess behavior. Years of saved time are no consolation for invalid results.

References
Ebbesen, E. B. & Allen, R. B. Further evidence concerning Fiske's question: 'Can personality constructs ever be validated? ' Unpublished manuscript, 1977. (Available from E. B. Ebbesen, Department of Psychology, University of California at San Diego, La Jolla, Calif. 92037.)
THE THEORY OF COGNITIVE ORIENTATION: WIDENING THE SCOPE OF BEHAVIOR PREDICTION

Hans Kreibig and Shulamith Kreibig

DEPARTMENT OF PSYCHOLOGY
TEL AVIV UNIVERSITY
RAMAT AVIV
TEL AVIV, ISRAEL

I. Introduction .......................................................... 102
II. Cognitive Orientation Theory: The Submodal Level ........ 103
   A. Meaning and Orientation ..................................... 103
   B. The Lowest Level of Cognitive Functioning .............. 107
   C. Primary Input Elaboration and Output Selection ......... 108
III. Cognitive Orientation Theory: Determinants of Molar Behavior .... 110
   A. Meaning Generation .......................................... 110
   B. Beliefs, Orientativenss, and the Four Cognitive Orientation Components .... 111
   C. Belief Evocation and Cognitive Orientation Clustering .... 114
   D. Behavioral Intent and Behavioral Program .............. 118
   E. Cognitive Orientation Theory and Personality Research ...... 121
IV. Some Elements of Technique .................................... 127
V. Brief Review of Former Studies .................................. 129
   A. Prediction Studies ........................................... 129
   B. Changing Behavior Studies ............................ 132
VI. Replications of Former Studies ................................. 133
   A. Predicting Pain Tolerance from Cognitive Orientation Components .... 133
   B. Predicting Punctuality from Cognitive Orientation Components .... 133
   C. Predicting Achievement Behavior from Cognitive Orientation Scores .... 136
VII. Predicting Behavior in Normals: New Studies ............... 137
   A. Predicting Behavioral Responses to Stress ................ 137
   B. Predicting Ordinalness from Cognitive Orientation Components .... 140
   C. Predicting Assertiveness from Cognitive Orientation Components .... 141
   D. Predicting Conformity from Cognitive Orientation Components .... 142
   E. Predicting Graduation Grades in Cadets’ School on the Basis of Cognitive Orientation Scores .... 141
   F. Predicting Social Withdrawal on the Basis of Cognitive Orientation Scores .... 145
VIII. Studies in Psychopathology .................................... 147
   A. Predicting Shyness in Mental Patients .................... 147
   B. Predicting Responses to Conflict in Schizophrenics and Normals .... 150
   C. Cognitive Orientation Components in Cognitive Representations .... 153