How Relevant is a Semantic Similarity Interpretation of Personality Ratings?

Jack Block, Daniel S. Weiss, and Avril Thorne
University of California, Berkeley

In recent years, a conceptual similarity interpretation of personality and interpersonal assessment ratings has been advanced by D'Andrade and Shweder. According to their view, such ratings are primarily understandable as linguistic artifacts having little or no connection with the real world. This position contravenes a central and widely used methodology of personality and behavior assessment. The present article critically evaluates the logic and data that have been used to support this conceptual similarity explanation and concludes that pertinent evidence is wanting. A primary finding of D'Andrade and Shweder, that judges can predict the intercorrelations among personality ratings, is shown in an experimental study to be an adventitious function of previously unrecognized patterns of redundancy in the personality variables that happen to have been used. Finally, some empirical relationships not assimilable to their position are noted together with some conceptual problems of their viewpoint.

A decade ago, D'Andrade (1965, 1974) began to advance a radical and sweeping critique of two essential techniques of personality assessment: personality ratings and personality questionnaires. Applying his own analytical logic to several sets of previously published data, D'Andrade concluded that the correlations [among rating and inventory scales] are primarily an artifact of the rater's or the questionnaire taker's cognitive structure, and not a reflection of the real world (1974, p. 181).

More recently, Shweder (1975, 1977a, 1977b) has extended and elaborated on D'Andrade's theme, asking the confronting question: "How relevant is an individual difference theory of personality?" (1975, p. 355). By "an individual difference theory of personality," Shweder referred to generally held conceptions of personality as epitomized, for example, by Child's (1968) definition: "more or less stable internal factors that make the person's behavior consistent from one time to another, and different from the behavior other people would manifest in comparable situations" (p. 83). Shweder answered his question negatively, boldly asserting that "an individual difference theory of personality" has been "shown" by him to be "no more than statements about how respondents [and psychologists] classify things as alike in meaning" (1975, p. 482). As Shweder noted, much of the evidence for the existence of personality parameters that shape behavior in coherent ways has depended on observation-based ratings and personality inventories. According to Shweder, when rating and inventory data are generated by psychologist-observers or by inventory takers, respondents (including psychologists)

unwittingly substitute a theory of conceptual likeness for descriptions of behavioral co-occurrences. . . . Items alike in concept are inferred to be behaviorally characteristic of the same person even when, as is typically the case, conceptual relationships among items do not correspond to the actual behavioral relationships among items. . . . [As a result] these conceptually biased judgments create an "illusion" of underlying behavioral consistency which, although not apparent in actual behavior, deceptively validates the "individual difference" conceptualization of "personality." (1975, pp. 455-456)

This point of view, which strikes at the

This study was supported by National Institute of Mental Health Grant SRO MH 16080 to Jack Block. Requests for reprints should be sent to Jack Block, Department of Psychology, University of California, Berkeley, California 94720.
very core of personality and behavioral assessment, has recently achieved appreciable currency, being approvingly cited in reviews by Schneider (1972) and Mischel (1968, 1969, 1971, 1977), among others. The present article is an effort to evaluate the premises, logic, and empiricism that underlie the strong conclusions of what we shall call the conceptual similarity argument. We focus primarily on the "reference" of personality ratings, a quite sufficient context in which to address the issues that have been raised.

The Conceptual Similarity Argument

Underlying the conceptual similarity position are several propositions and presumptions about personality assessment that require specification and scrutiny if the tenability of this conclusion is to be evaluated.

The Conceptual Similarity View of Personality and Behavior Assessments: Emphasis on Trait Similarity Structures

D'Andrade characterizes personality assessment as follows:

First, one or more human observers are asked to judge one or more subjects on a number of traits of behavior; the judgments are expressed in ratings or rankings (scores) based on the observers' long-term (e.g., more than 30 minutes) memory of the subject's behavior. Second, a single score is computed for each subject for each trait, usually by taking the mean of all the scores given each subject for each trait. Third, to find out how the traits are related to each other, some measure of association, such as the product-moment correlation, is computed from the subjects' scores for all pairs of traits. Finally, the measures of association are analyzed to determine how the traits are organized with respect to each other. The results of such analyses indicate which traits tend to go together, and the similarity structure of the trait measurements is taken as a representation of the structure or organization of the subject's behavior. (1974, p. 162)

This impression of personality and behavioral assessment will appear foreign to most assessment psychologists. Insofar as ratings are used as one of the primary methods of assessment, the first two steps listed by D'Andrade often apply. The rating data thus obtained, however, are then generally used in a multiphased effort (together with other kinds of measures) to establish the construct validity of a concept. The goal of such effort is a dependable and coherent nomological network of relationships on which to base subsequent theorizing or prediction. The emphasis of personality research is, centrally, on the validity of the measures being used. That is, do individuals earning different scores on an index also earn conceptually testable scores on the index? We recognize three independently formulated measures. There is a long history to this kind of effort; recent reviews of the status, accomplishments, and deficiencies of personality research are to be found in Block (1976), Gough (1976), Hogan, DeSoto, and Solano (1977), and Stagner (1977).

The third and last step listed in D'Andrade's (1974) conceptualization of personality research may—often do, indeed—be an illusion when multiple ratings or measures have not been gathered and there is question or concern about the extent of redundancy of coverage of the personality or behavioral domain by the variables used. Although a certain amount of redundancy is necessary to ensure reliability and validity beyond this amount is wasteful. Various clustering methods, such as factor analysis, can be extremely useful in quickly and simply revealing the latent dimensions of the set of ratings employed and as a guide to identifying rating variables that are dispensable with little, or at least acceptable cost.

Although the results of factor analyses (and related methods) of sets of ratings are often immediate and intrinsic interest, there are few personality psychologists today who would claim that obtained factor structures represent inexorable verities (cf. Lykken, 1971). It is now well recognized that the degree of conceptual similarity or meaning equivalence each variable has with respect to every other variable. The resulting matrix of similarity indices we call a Conceptual Similarity matrix. The primary demonstration of the conceptual similarity position has been to show that for a number of previously published Rated Behavior matrices, the pattern of intercorrelations characterizing a Rated Behavior matrix could be well approximated by the pattern of intercorrelations characterizing the Conceptual Similarity matrix based on the same set of variables. Because of the frequently strong correspondence between these two matrices, in which the Rated Behavior matrix is a different construction for each individual, whereas the Conceptual Similarity matrix derives from purely semantic judgments of the similarity of personality variables, made without reference to palpable people, D'Andrade concluded that "with respect to personality and behavior specifications in the field of psychology . . . it is possible to confine propositions about the world with propositions about language." (1965, p. 215). Sweder has been more vigorous in his assertions, claiming that evidence generally thought to demonstrate the applicability of the "individual difference" conceptualization of "personality" has been shown to be exemplified by deriving the trait-factors discovered in interpersonal rating forms, questionnaire interviews and personality inventories from purely conceptual criteria without any reference to actual behavior. (1975, p. 493).

We suggest that the correspondence between a Rated Behavior matrix and a Conceptual Similarity matrix, when it exists, does not have the logical implication with respect to individual differences that the conceptual similarity position seems to claim. The reader should pause for a moment to verify that in no way whatever is it possible to proceed from the specific similarity judgments on which the Conceptual Similarity matrix is based to a specification of just which variables are rated high or low on a particular rating dimension. For example, to know that judges evaluate the personality rating variables "Introverted" and "Devaluating oneself" as conceptually highly similar (Sweder, 1975, Table 1) is to be entirely uninformative about who is rated Introverted or as Devaluating oneself or whether Introverted and Devaluating oneself are separate or combined variables, important and useful characteristics in terms of which to classify people. We shall return later to this distinction between classifying personality variables as opposed to classifying individuals in terms of personality variables.

The Secondary Demonstration of the Conceptual Similarity Position: Absence of Correspondence Between Rated Behavior and Actual Behavior Matrices

From the beginning, D'Andrade recognized the confounded nature of the primary demonstration: it is possible that the so-called psychological traits
Later in this article, we shall suggest a number of reasons why the studies located by D'Andrade and Shweder fail to meet their special hypothesis-testing requirements.

The Implications of the Conceptual Similarity Position

The critical import of the conceptual similarity argument is powerful, blunt, and devastating to personality psychology and other behavioral science fields relying on some form of ratings.

If the correlations on which these studies rest are primarily an artifact of the raters or the questions taken's cognitive structure, and not a function of the real world, there is little or no evidence that human behavior can be described by maladaptive units. What remains is a world in which human behavior is to be described in terms of specific behaviors occurring in specific situations as Much and others have argued. (D'Andrade, 1974, p. 183)

D'Andrade went on to suggest that it is not necessary to group behavior into clusters or traits, or dimensions to be able to give an economical description of an individual's behavior. Instead, he claimed that a remarkably predictive description can be obtained (p. 183) if the simplest model of the most frequent behavior previously observed in each of a number of different behaviors are considered with regard to the predictive efficacy of this behavior-specific approach. Whatever the predictive accuracy of this actuarial orientation may prove to be, it is clear that it abandons the elegance and economy of a theory of behavior. Many psychologists will not be interested in a psychology reduced to, and aspiring to no more than, an endless graphing of frequency distributions for behavior-category systems. Were the logic and epistemology underlying the behavior-specific approach compelling, this reductionistic but assumption-free approach to "good prediction" would prevail. It is our contention, however, that the argument and evidence brought together to buttress the conceptual-similarity interpretation of personality ratings as existing only "in the eyes of the beholder" are deeply flawed, wrongly focused, and greatly embarrassed by a host of empirical findings unsatisfactory to the systematic-distortion position.

The remainder of this article will more closely consider the evidence and reasoning underlying the conceptual similarity interpretation of the processes underlying personality assessment. We shall first evaluate the Actual Behavior -- Rated Behavior matrix analyses intended to confound the isomorphic and systematic-distortion hypotheses. Next, we go on to illustrate the conditions that are sufficient and perhaps necessary to create or to prevent correspondence between Rated Behavior and Conceptual Similarity matrices. Finally, we list a variety of connections between personality ratings and incommensurate real world outcomes that cannot be understood in artificial terms. In the course of our counter-evaluation, we shall burden the reader with matters of analytic detail, psychometrics, and principles of inference. We do not apologize for the lengthiness of our analysis. Too often, human psychology, crucial experimental points, and analytical considerations go unnoticed. If the reader is to form well-based opinions on the important issues involved, he or she will have to be actively involved in the specifics of the evaluative process.

Unconfounding the Isomorphic and Systematic-Distortion Hypotheses: Some Problems in the Analyses

To harden the case for a conceptual similarity interpretation of personality ratings, demonstrations were sought that:

- Verbal data collecting procedures, such as interpersonal ratings, interview protocols, and retrospective questionnaires, cannot be systematically supported by detailed and immediately recorded observational data. (Shweder, 1974, pp. 464-466)

It was reasoned that patterns of behavior (as Actual Behavior matrices) with Rated Behavior matrices) is needed here because it is not the validity of specific traits or categories which is in question (emphasis added), but the validity of the correlations between traits. Ideally, for recording the actual behavior of the subjects, a mechanical device should be constructed to count frequencies of different kinds of behavior. Unfortunately, because a mechanical measuring device is not practical at present for any judgment more complex than is making some versus is not making some simplification through less exacting observational devices appears to be a trained observer using a simple coding scheme to record a subject's behavior as it occurs. The trajectory of an observer's assessment and the simplicity of the coding decisions should, it is hoped, predict against systematic distortion of the type thought to take place in long-term (e.g., over a ten minute) memory. (D'Andrade, 1974, p. 162)

The two studies reworked by D'Andrade for his Actual Behavior -- Rated Behavior matrix comparisons are an investigation by Borgatta, Cottrell, and Mann (1958) and the doctoral dissertation of Mann (1959). Shweder has added a comparable analysis of a study by Newcomb (1929). It will be helpful to briefly characterize these studies.

The Borgatta et al. (1958) research used 47 graduates divided into five small groups. Each group met for two-hour sessions to discuss "democratic leadership." During the ninth and tenth sessions, a single observer using the Interaction Process Analysis (IPA) category system (Bales, 1961) evaluated the ongoing interaction. Subsequently, for each member of the group, the rate of each Bales category was calculated. These rates served as the Actual Behavior scores for subsequent analysis. After the ninth session, each group member ranked all the members of his group with respect to 40 traits, 6 of which approximated the Bales categories. Subsequently, for each member, the average ranking (adjusted for the slight differences in group sizes) on each of the 6 traits approximating the Bales categories was calculated. These average rankings served as the Rated Behavior scores for subsequent analysis.

The Mann (1959) study used 100 fraternity undergraduates, each of whom was assigned to different and nonoverlapping five-man groups. In one of these 50-minute groups, the group worked on a specific task; in the other kind of group, the group members attempted to formulate some fraternity policies. For each of the first group, a single observer used a version of the Bales interaction categories to code the ongoing interaction. For each member of the group, the total
number of "acts" was calculated. For each of
seven Bales IPA categories, the subject's act
frequency was divided by the total number
of acts to derive a percentage. These per-
centages served as the Actual Behavior scores
for subsequent analysis. After a group ses-
sion, each member rated the other four on
16 characteristics, 7 of which approximated
the Bales-Mann interaction categories. Sub-
sequently, for each subject, the sum of the
ratings assigned by the other four members
of his group on each of the characteristics
was calculated. These pooled ratings served
as one form of Rated Behavior scores for
subsequent analysis. In addition, the inter-
action-observing coding at the end of each
session was read directly for each
subject the percentage of the subject's be-
havior in each of the Bales-Mann categories.
These estimated percentages served as a sec-
ond form of Rated Behavior scores in sub-
sequent analysis.

The Newcomb (1929) study evaluated 21
problem boys (a group of 21 preadolescents
attending a month-long summer camp) for each
group. Each group consisted of six tentfuls of boys, each
tent supervised by a different counselor.
Sometimes during each day, for each boy, his
own counselor completed a behavior record
with respect to 26 behavioral questions New-
comb asked because of their presumed rel-
ance to extraversion-introversion. For each
question, the counselor selected one of four
alternatives provided by Newcomb, coded as
"positive" (i.e., implying extraversion) or
"negative" (i.e., implying introversion). In
subsequent tabulation for each boy, for each
of the behavior questions the number of days
on which a positive or negative alternative
had been recorded was divided by the total
number of camp days for which a report had
been submitted. These 26 proportions, from
table and uncalibrated observers, served as
the Actual Behavior scores for subsequent
analysis. At the end of the camp period, a
rating form on approximately the same 26
behavior situations was completed for each
boy by his tent counselor and by five other
counselors who had frequently observed each
boy. The six ratings for each boy were then
combined into one consensus rating. The 26
pooled ratings for each boy served as the
Rated Behavior scores for subsequent ana-
lyses.

D'Andrade (1974) and Swedler (1971)
suggest that the three studies just described
permitted them to "explicitly separate the
structure of social impressions from the
structure of behavioral instances" (Swedler,
1971, p. 479). We suggest, alternatively, that
the reasonably and evident independence of
D'Andrade and Swedler as support for the
"separation" lacks power and pertinence.
Some of our concerns follow.

The categories or items or dimensions of
behavior measured by the Actual Behavior
indices are often different from those rated
by observers. An analysis seeking to inter-
pret the differences between the similarity
structures derived from an Actual Behavior
and a Rated Behavior matrix in conceptually
similar terms requires that the sets of
variables be considered as distinct, of course
be identical. The only data difference should
be that in the Actual Behavior case, the vari-
able is directly from "actual behavior," whereas in the Rated Behavior case the vari-
able is scored on the basis of retrospective
judgments and.inferences.

However, in none of the three studies did
the differences between the actual cases and
the rated cases clearly absolve that variables
were scored on the basis of retrospective
judgments and inferences.

In the Borgatta et al. (1953) study, the
actual behavior variable, IPA Category 2,
defined as "Shows solidarity, raises others'
status, takes, gives help, reward." The corre-
responding Rated Behavior variable is de-
finied as "Shows solidarity and friendship." Such facets of interaction as raising others'
status, joking, and giving help and reward
included in the Actual Behavior behavior of
this variable, are not included in the Rated
Behavior version. The Actual Behavior vari-
able, IPA Category 2, is defined as "Shows
tension release, shows satisfaction, laughs." The corresponding Rated Behavior variable
is defined as "Is responsive to laughter.
Such facets of interaction as showing tension
release and showing satisfaction, included in
the Actual Behavior version, are omitted
from the Rated Behavior version. The Ac-
tual Behavior variable, IPA Category 3, is
defined as "Gives suggestion, direction, im-
plying autonomy for other." The correspon-
ding Rated Behavior variable is defined as
"Makes the most suggestions." The impor-
tant qualifier, "implying autonomy for oth-
er," present in the Actual Behavior ver-
sion, is absent from the Rated Behavior ver-
sion. The remaining Actual Behavior - Rated
Behavior correspondences pose similar prob-
lems of comparability.

In the Mann (1932) study, the various Bales
categories were merged. The first Actual Be-
havior variable, Mann's Category 1, sub-
sumes showing solidarity (excluding all jok-
ing behavior) and showing agreement. The
Rated Behavior version of this variable in-
quired "how much [the group member being
rated] tended to agree with what others had
said." No mention of solidarity is contained
in the Rated Behavior variable definition, a
most important omission.

Swedler's (1975) report of the Newcomb
(1929) study claims that the same 26 items
of behavior were evaluated by the daily
records and by the end-of-camp ratings. In
fact, 15 of the 26 items lack Actual Be-
"behavior - Rated Behavior identity. For
example, the daily record item, "Was he fond
of swimming?" becomes the rating item,"Did
he spend his swimming periods in the water,
actively moving about?" The daily record item,
"How much of the day did he spend doing things that required little or
no action?" becomes the rating item,"Was
he actively moving about most of the day?"
Sometimes these differences seem slight but,
as is well known, slight differences in phrasing often can have great effect.

The absence of exact equivalence between
the definitions of the Actual Behavior and
the Rated Behavior variable sets is regret-
table, this consideration alone renders prob-
lematic the significance of subsequent Actual
Behavior - Rated Behavior matrix compar-
isons that D'Andrade and Swedler have
been crucial for the essential in-
validity of their conclusions.

Emphasis on the correlations be-
tween traits ignores consideration of the
validity of specific traits. We believe that
most psychological traits are not a single
trait, but extend into a family of traits.
Only a detailed analysis of the trait famil-
ies can lead to a full understanding of the
trait structure. A simple analysis of trait
validity alone is likely to lead to misun-
derstandings. This is the approach that
D'Andrade and Swedler have taken, but
the approach is not appropriate for these
data. The question of trait validity should
be considered in the context of the full
trait structure. The validity of a trait fam-
ily is complex and cannot be assessed
simply by examining the correlation be-
tween two traits. The validity of a trait
family is a matter of the structure of the
trait family as a whole.
The use of frequency counts of behavior is not a sufficient means of operationalizing complex psychological concepts. Conceptual similarity advocates recommend, "for recording the actual behavior of the subjects, . . . count frequencies of different kinds of behavior . . . using a simple coding scheme." (D'Andrade, 1974, p. 162). The several assumptions underlying this epistemological position are bothersome on several grounds.

1. D'Andrade and Shweder write if "actual behavior" exists; it does not. Behavior in the raw has an infinite variety of facets and pattern possibilities and therefore can only be studied by selection. The act of selection, however, done, represents a constructive and theoretical assertion about the world that was its justification by the (ultimately personal) nomological network with which it subsequently can be surrounded. Kaplan (1984) has written telling regarding the philosophical issues in making the decision as to what to observe. More recently, Neisser (1976) has summarized various lines of evidence from cognitive psychology that indicate the human organism is an active, schematizing but also schematizing-modifying being who only selectively connects with uncomparably "real.

It should also be noted that the problems surrounding the notion of "actual behavior" cannot be escaped by shifting to a focus on "objective behavior." So-called "objective" measures of behavior of course can only reflect earlier decisions, sometimes not fully considered, as to where attention should be addressed. A conceptually inappropriate decision as to what aspect of behavior to select as an index of a psychological dimension cannot be redeemed by the subsequent partiality of the coding of the irrelevant behavior. The scientific merit of "objective behavior" depends, finally, on the psychological correctness of what is being figured from the ground, not on the mechanicalness of recording.

What D'Andrade and Shweder appear to mean, then, by "actual behavior" are frequency counts or proportions summarizing arbitrary and unspecified observations from the infinitely broad stream of behavior, observed by an individual who need make no inference and who records the single-out behaviors immediately. Such counts or proportions are viewed as an ultimate or at least sufficient criterion of truth—being the benchmark against which all other data can be referenced and tested. However, we suggest that large problems assail even this more restricted view.

2. It is assumed that the Actual Behavior data are simple and direct in two senses: No pertinent behaviors are omitted, and no inference is required of the recording observer. Both of these assumptions are questionable. Interaction Process Analysis, used to provide Actual Behavior data in both of the studies offered by D'Andrade (1965, 1974), has been found to have many problems and is no longer viewed with the enthusiasm that surrounded it when first introduced (see Longabaugh, 1963; O'Dell, 1968; Waxler & Meshler, 1966). IPA coding is a taxing and inducting task: The solitary observer records "acts" continuously, typically at a rate of 10 to 20 scores per minute (Scales, 1965) or 50 minutes (in the Mann, 1959, study) or for 120 minutes (in the Bogartta et al., 1958, study), coding the ongoing stream of "acts" of group members into 12 supposedly sufficient, mutually exclusive, and by no means simple categories (7 categories in the Mann study) according to the particular
became a rejected isolate by the end. The behavior percentage scores used to reference "actual behavior" are highly reliable and cannot index such changes. On the other hand, the "rated behaviors" in the Newcomb study have the possibility of expressing such changes as may have occurred. Making their ratings at the end of the camp session, the raters had the opportunity to express their integration of all of their understandings of each boy, including such contextual or sequential or salient effects as they believed to be pertinent. In short, useful though the average basis underlying the "actual behavior" scores may generally be, such scores may not be able to capture important information and recognitions codable by raters who use long-term memory. Besides the possibility of "systematic distortion" emphasized by D'Andrade (and Shwedler) when raters invoke long-term memory, there is also the possibility that raters will be able to discern relationship, meanings, and trends that were denied identifiability in the highly summarizing averages. Again, we suggest that most personality psychologists will not accept such averages uncritically as a criterion of "actual behavior."

Extraneous influences distorting correlation coefficients; and their presence are not considered. Many factors influence individual correlation coefficients and, thus, the subsequently organized matrix of correlation coefficients. Sometimes these correlation-influencing factors are not of great importance vis-à-vis the broad questions being asked of the data. But when conclusions hinge on the particular values of particular pattern-making or pattern-breaking coefficients, these considerations become crucial to evaluate and to control. Yet, when D'Andrade and Shwedler contrast matrices, they ignore the many psychometric and statistical considerations that influence the observed differences in correlational patterning.

The distributions of the Actual Behavior scores and the Rated Behavior scores are not considered. They are often likely to be highly skewed or not unimodal, particularly the IPA rates or percentages representing "actual behavior." Correlations between variables are importantly influenced (and always lowered) by differences in the shape of the distributions being related (Carroll, 1964).

Furthermore, these 15 or 21 "observations" are not independent of each other; the correlation between Categories X and Y and the correlation between Categories X and Z will constrain the correlation between Categories Y and Z. The distribution of independent observations will tend to be strange and unstable, reacting strongly to changes in the constitution of the set of five or seven categories that happen to have been used. Correlations based on nonindependent observations are not correlations in the usual sense of the term, since the interdependency of observations violates a fundamental assumption of the statistical logic underlying the correlation coefficient. One can no longer interpret such fluky coefficients in terms of amount, of variance explained or significance level. The correlation between two correlation matrices has a certain descriptive value, but the intuitive appreciation of these coefficients requires the further recognition that these coefficients are indeed of an unexplainable but certainly bizarre metric.

The direction of scoring or rating a psychological variable is usually arbitrary; one can algebraically reflect a variable without changing its conceptual import. Instead of assigning high numbers to indicate "extroversion," we could as well assign high numbers to mean "introversion." This long-established recognition has severe implications for the "correlation" between two matrices.

Changes in the direction in which a variable is scored reverse the sign of the correlation of that variable with other variables and will fundamentally influence the "correlation" subsequently computed between these revised matrices (Tellegen, 1963). In the D'Andrade and Shwedler analyses, bipolarities tended to characterize their redundancies of measurement. Had the variables defining one end of their dimensions been defined instead in reflected terms, the subsequent correlation matrix would have had fewer negative entries and the "correlation" between these variables would have been importantly smaller. Thus, in yet another way, the results of these analyses are shaped by adventitious aspects of the data.
in the particular variable sets D'Andrade and Shwedder sampled.

By varying the amount and structure of variable redundancy, thus influencing the extent and pattern of intercorrelations among variables, we have been able to manipulate the subsequent degree of correspondence between Rated Behavior and Conceptual Similarity matrices. Our logical demonstration, together with the recognition of the kinds of redundancy present in the variable sets evaluated by D'Andrade and Shwedder, diminishes the implications of their findings.

Method

A Demonstration of the Working of Redundancy

For our manipulative purposes, a set of personality variables was required from which sub-sets could be selected that met specified redundancy designs. The California Q-set (CQ-set; Block, 1951), a broadly ranging set of personality-descriptive items, was used because of its extensive use in personality assessment and because of its multidimensional characteristics. The CQ-set had been used as a basic personality description procedure in an longitudinal study of personality development (Block, 1951). In this study, the personality characteristics of 871 men were assessed by means of the CQ-set at nine time periods: junior high school, senior high school, and adulthood (at the ages of 40 and 45).

The results indicated that a set of 12 items, representing two factors each, were selected to represent the first, second, and fourth orthogonal factors. For each factor, four CQ-items were included: two loaded highly on the first factor and two loaded highly on the second factor. The average factor loading of these 12 items (6-8 for each factor) was 0.39. These 12 CQ-items were selected to represent a multidimensional and bipolar redundancy and will hereafter be called the Unidimensional Bipolar Redundancy variable set. The 64 intercorrelations among these 12 variables constituting the Unidimensional Bipolar Redundancy variable set were calculated.

The second set of 12 CQ-items was chosen to represent the first, second, and fourth orthogonal factors. For each factor, four CQ-items were included: two loaded highly on the first factor and two loaded highly on the second factor. The average factor loading of these 12 items (6-8 for each factor) was 0.39. These 12 CQ-items were selected to represent a multidimensional and bipolar redundancy and will hereafter be called the Multidimensional Bipolar Redundancy variable set. The 64 intercorrelations among these 12 variables constituting the Multidimensional Bipolar Redundancy variable set were calculated.

The third set of 12 CQ-items was chosen so that each item represented 1 of 12 orthogonal factors. Some of the latter factors were extracted by only a few CQ-items. In selecting an item to represent a factor, some attention was paid to the item's loadings on the other 11 factors as well as the item's orthogonality to the other factors. The average factor loading of these 12 items (6-8 for each factor) was 0.39. These 12 CQ-items were selected to represent a multidimensional and bipolar redundancy and will hereafter be called the Multidimensional Bipolar Redundancy variable set. The 64 intercorrelations among these 12 variables constituting the Multidimensional Bipolar Redundancy variable set were calculated.

The results indicated that a set of 12 items, representing two factors each, were selected to represent the first, second, and fourth orthogonal factors. For each factor, four CQ-items were included: two loaded highly on the first factor and two loaded highly on the second factor. The average factor loading of these 12 items (6-8 for each factor) was 0.39. These 12 CQ-items were selected to represent a multidimensional and bipolar redundancy and will hereafter be called the Multidimensional Bipolar Redundancy variable set. The 64 intercorrelations among these 12 variables constituting the Multidimensional Bipolar Redundancy variable set were calculated.

Table 1: California Q-Set Items Constituting the Three Variable Sets

<table>
<thead>
<tr>
<th>Items constituting the Unidimensional, Bipolar Redundancy set</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Is a talkative individual.</td>
</tr>
<tr>
<td>17. Has a rapid personal tempo; behaves and acts quickly.</td>
</tr>
<tr>
<td>20. Is facetly and / or garrulous expressive.</td>
</tr>
<tr>
<td>23. Behaves in an assertive fashion in interpersonal situations.</td>
</tr>
<tr>
<td>24. Tends to be overly negative about self or any other person.</td>
</tr>
<tr>
<td>33. Is self-dramatizing; histrionic.</td>
</tr>
<tr>
<td>34. Tends to be overly negative about self or any other person.</td>
</tr>
<tr>
<td>45. Alas, keeps people at a distance; avoids close interpersonal relationships.</td>
</tr>
<tr>
<td>61. Is emotionally bland; has flattened affect.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Items constituting the Multidimensional, Bipolar Redundancy set</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Is a talkative individual.</td>
</tr>
<tr>
<td>17. Has a rapid personal tempo; behaves and acts quickly.</td>
</tr>
<tr>
<td>20. Is facetly and / or garrulous expressive.</td>
</tr>
<tr>
<td>23. Behaves in an assertive fashion in interpersonal situations.</td>
</tr>
<tr>
<td>24. Tends to be overly negative about self or any other person.</td>
</tr>
<tr>
<td>33. Is self-dramatizing; histrionic.</td>
</tr>
<tr>
<td>34. Tends to be overly negative about self or any other person.</td>
</tr>
<tr>
<td>45. Alas, keeps people at a distance; avoids close interpersonal relationships.</td>
</tr>
<tr>
<td>61. Is emotionally bland; has flattened affect.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Items constituting the Approximately Orthogonal set</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Appears to have a high degree of intellectual capacity (whether actualized or not). Orientability is not necessarily assumed.</td>
</tr>
<tr>
<td>8. Is thin-skinned; vulnerable to anything that can be construed as criticism or an interpersonal slight.</td>
</tr>
<tr>
<td>17. Has a rapid personal tempo; behaves and acts quickly.</td>
</tr>
<tr>
<td>20. Is facetly and / or garrulous expressive.</td>
</tr>
<tr>
<td>23. Behaves in an assertive fashion in interpersonal situations.</td>
</tr>
<tr>
<td>24. Tends to be overly negative about self or any other person.</td>
</tr>
<tr>
<td>33. Is self-dramatizing; histrionic.</td>
</tr>
<tr>
<td>34. Tends to be overly negative about self or any other person.</td>
</tr>
<tr>
<td>45. Alas, keeps people at a distance; avoids close interpersonal relationships.</td>
</tr>
<tr>
<td>61. Is emotionally bland; has flattened affect.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Items constituting the Multidimensional, Bipolar Redundancy set</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Is a talkative individual.</td>
</tr>
<tr>
<td>17. Has a rapid personal tempo; behaves and acts quickly.</td>
</tr>
<tr>
<td>20. Is facetly and / or garrulous expressive.</td>
</tr>
<tr>
<td>23. Behaves in an assertive fashion in interpersonal situations.</td>
</tr>
<tr>
<td>24. Tends to be overly negative about self or any other person.</td>
</tr>
<tr>
<td>33. Is self-dramatizing; histrionic.</td>
</tr>
<tr>
<td>34. Tends to be overly negative about self or any other person.</td>
</tr>
<tr>
<td>45. Alas, keeps people at a distance; avoids close interpersonal relationships.</td>
</tr>
<tr>
<td>61. Is emotionally bland; has flattened affect.</td>
</tr>
</tbody>
</table>

The purpose of this research is to obtain judgements of the similarity between self-descriptions. Each pair of descriptions is on a card. To familiarize yourself with the cards, spread the cards out and shuffle them. Then, taking the cards in any order you wish, read the ten descriptions on a card and decide how conceptually alike are the two self-descriptions. Make each judgement according to the following scale, placing each card in the pile corresponding to your judgment: 7 = extremely similar; 6 = moderately similar.
similarity; $= mildly similar; 4 = unrelated, neither
similar nor dissimilar; $= mildly dissimilar; 2 =
moderately dissimilar; 1 = markedly dissimilar.

These instructions were adapted from those previously
employed by D'Andrade and Shwedler.

For the Unidimensional, Bipolar Redundancy variable set and the Multidimensional, Bipolar
Redundancy variable set, 10 judges each made simi-
larity judgments. For the Approximately Orthogonal
set, 20 judges were used. No judge rated more than
one variable set. The judges were undergraduates
in a course on the psychology of personality and
were uninformed as to both the nature and the
purpose of the study. Two of the 20 judges of the
Approximately Orthogonal variable set did not cor-
rectly complete the task, reducing this final total to
18.

Results

The 66 judgments of conceptual similarity by a judge is his or her Conceptual Similarity
matrix. When the individual Conceptual Similarity matrices of the Unidimensional, Bipolar Redundancy judges are correlated with the six Unidimensional, Bipolar Redundancy Rated Behavior matrices, the average Conceptual Similarity - Rated Behavior In-
ternatrix “correlation” is .62, ranging from .10 to .95. When the individual Conceptual Similarity matrices of the Multidimensional, Bipolar Redundancy judges are correlated with the six Multidimensional, Bipolar Re-
dundancy Rated Behavior matrices, the average Conceptual Similarity - Rated Behavior In-
ternatrix “correlation” is .50, ranging from .31 to .75. When the individual Conceptual Similarity matrices of the Approximately Orthogonal judges are correlated with the six Approximately Orthogonal Rated Behavior
matrices, the average Conceptual Similarity - Rated Behavior Internatrix “correlation” is .41, ranging from .10 to .46. As expected, the mean Conceptual Similarity - Rated Behavior “correlations” of the Unidi-
menional, Bipolar Redundancy and Multidi-
menional, Bipolar Redundancy groups are not significantly different ($t = .99, p < .38$). When the Unidimensional, Bipolar Redundancy and Multidimensional, Bipolar Redundancy judging groups are combined and con-
trasted with the Approximately Orthogonal judging group, the difference between their respective mean Conceptual Similarity -

Rated Behavior Internatrix “correlation” is highly
significant ($t = 5.17$, $p < .0001$). Thus, these data reveal, strikingly, that Conceptual Similarity - Rated Behavior Internatrix “correlation” is high when bipro-
redundancy is high and is low when biproredu-

redundancy is low. Judges appear to rely rather
faithfully reflectors of the pattern of redundancy empirically existing within the set of variables they are evaluating for Conceptual Similarity.

As a related finding, the average Internatrix “correlation” among the Conceptual Similarity
matrices contributed by the 10 Unidimen-

dimensional, Bipolar Redundancy judges was .41. The average “intercorrelation” among the Conceptual Similarity matrices contributed by the 10 Multidimensional, Bipolar Redundancy judges was .50. The average “intercorrelation” among the Conceptual Similarity matrices contributed by the 10 Approximately Orthogonal judges was .21. Thus, it appears that the

availability of judges to agree among themselves

in their judgments of conceptual similarity is a clear function of the pattern of redundancy present among the variables being evaluated.

Redundancy in the Variable Sets

Evaluat ed by D’Andrade and Shwedler

The pertinence of our demonstration of the

influence of redundancy on the “correla-
tion” between Rated Behavior and Conceptual

Similarity matrices depends on the extent to

which the amount and pattern of redundancy, previously unrecognized inoperative, characterizes the variable sets sub-

sequently found to be “retrievable” by Con-

ceptual Similarity judgments.

Inspecting the variable sets that have been

used, it is clear, we suggest, that appreciated

and highly patterned redundancy demand

characterize them.

The 20 personality terms of Norman

(1968), further analyzed by D’Andrade

(1963), were a carefully refined and test-

ed.

*1 should be noted that the generally small but
constant leading factors of the Approximately Orthogonal

variables on the first factor contribute to the mean

Conceptual Similarity - Rated Behavior Internatrix

correlation of .21.

The studied variable set intended to provide re-

dundant specification of an orthogonal five-

factor structure deemed conceptually suitable

for encompassing the nominal data. Our Multi-

dimensional, Bipolar variable set was designed to emulate the redundancy de-

sign built into the Norman terms.

The six rated variables in the Borgatta et

al. study were included in a factor analysis (1958, Table 3). Four of the variables

loaded highly on their first interpersonal

factor; the remaining two loaded highly on their

second factor. Our Multidimensional, Bipolar

variable set approximately represents the

redundancy characteristics of the

Borgatta et al. set of rated variables.

The six (or seven) rated variables in the

Mann (1959) study were not submitted to a factor analysis but, since they parallel
closely the variables used by Borgatta et al., it is probably safe to assume that the pat-

tern of redundancy characteristic of the

earlier study also characterizes the later study.

The 26 variables used by Newcomb

(1929) were not and cannot now be studied systematically with regard to their redundancy characteristics. However, because Newcomb selected his variables for their presumed rele-
vance to the single broad dimension of ex-

traversion–introversion, we suggest that appreci-
able redundancy of the unidimensional, bipolar type characterizes the Newcomb vari-

able set.

The three other variable sets whose ma-

trices are reported by Shwedler (1975, 1977) to be “retrievable” via conceptual similarity

judgments—those of Bales (1970), of Sears, Maccoby, and Levin (1957), and of Block

(1965)—all were factor analyzed, as re-

ported in their initial publications, and all manifested appreciable redundancy of the

multidimensional, bipolar, or unidimensional, bipolar type, according to the investigators’ measurement intentions.

Thus, bipolar redundancy within the vari-

able sets appears to be a strong and con-
stant feature of the D’Andrade and Shwedler ana-

lyses. This bipolar redundancy is generally

anomalous (and synonymous) in nature, thus

offering a semantic basis to judges for esti-

mating, if only approximately, the nonzero trajectories among various pairs of variables in the set. To the extent that Rated Be-

havior variables are constrained totally validly

with antonymous (and synonymous) redu-

dancy, the possibility of correspondence ("retrievability") between a Rated Behavior and a Conceptual Similarity matrix tends to disappear because the "similarity structure" of each matrix tends to disappear. The exter-

nal validity of the variables in the vari-

able set, either singly or as subsets, is in no way contingent on the "retrievability" or the absence of "retrievability" of a Rated Behavior matrix by a Conceptual Similarity matrix. To this last concern—the predictive and concurrent validity of personality rat-

ings—now turn.

Some Empirical Relationships Currently

Beyond the Reach of the Conceptual

Similarity Position

For assessment psychologists, a frustrating

aspect of the current controversy has been the

narrowly drawn arena in which the argument has been waged. In this section we introduce some relationships of a kind that personality psychologists would value of their belief in reliable individual differences with regard to personal parameters of consequence in the real world. We believe the kinds of across-time or across-domain rela-

tionships to be reported are either unassimil-

able to the conceptual similarity interpreta-

tion or else are encompassable only by a string of conjectures so complicated and ten-

uous as to deprive the conceptual similarity view of interest until such time as the necessary empirical support becomes available.

Our strategy for affirming the usefulness and

even the necessity of the conceptual–epi-


demic enterprise of personality requires even the necessity of the conceptual–em-

phasis on classifications of personality vari-

ables. We focus instead on the construct validity of classifications of individuals with respect to personality variables. Space limi-
tations permit only brief citation of some of the

available data confronting the conceptual similarity position.

The six Approximately Orthogonal Rated
Behavior matrices could not be well "retrieved" by the Approximately Orthogonal Conceptual Similarity matrix. By the reasoning of D'Andrade and Shevitz, it follows therefore that "an illusion of . . . behavioral consistency" (Shweder, 1975, p. 456) was not created by the conceptual similarity among the personality variables involved simply because there was not appreciable "conceptual likeness" among the variables being evaluated. If there is behavioral consistency nevertheless, such consistency is not illusory.

The "behavioral consistency" view or "individual difference theory of personality," as defined by Shweder (and as rejected by him), assumes that an individual's behavior will tend to be "consistent from one time to another, and different from the behavior other people would manifest." (Child, 1968, p. 83). A direct test of behavioral consistency is the correlation of a Rated Behavior variable as rated at one time with the same Rated Behavior variable as independently rated at another time.

For the 12 Approximately Orthogonal Rated Behavior variables, across-time correlations are available, for the sexes separately, connecting the junior high school (JHS) and senior high school (SHS) ratings (a period of 3 years), connecting the SHS and Adulthood ratings (a period of about 20 years), and connecting the JHS and Adulthood ratings (a period of about 25 years). It should be recalled again that these ratings were based on entirely independent sets of data and entirely independent sets of judges.

For the male and female samples from JHS to SHS, the 12 Approximately Orthogonal rated variables correlate, on the average, .42 and .40, respectively. If these correlations are adjusted for their attenuation due to unreliability, the mean correlations rise to .68 and .66. From SHS to Adulthood, the average correlations are .27 and .30 for the two sexes, respectively. Adjusted for attenuation, these mean correlations become .41 and .39. From JHS to Adulthood, the average correlations are .24 and .28 for the two sexes, respectively. Adjusted for attenuation, these mean correlations are estimated to be .36 and .46.

It should be noted that the 17 variables for which across-time person-ordering consistency is reported were not selected for consideration because of the across-time consistency they manifested. Many more such across-time correlations exist, the full set being reported by Block (1971). It should also be recognized that these across-time correlations are further and appreciably lowered by such developmental changes and transformations as take place from early to late adolescence and from late adolescence to the mid-thirties.

All things considered, we suggest that these across-time correlations are substantial evidence for the existence of sufficient behavioral consistency in sufficient people over sufficient time to warrant maintenance of the individual difference theory of personality. A close and empirically supported explanation of these and related data in conceptual similarity or other attributional terms has yet to appear.

In seeking to dispense with personality ratings and personality inventories on the basis of their conceptual similarity arguments and analyses, D'Andrade and Shevitz appear to have presumed the essential validity of personality assessment procedures. However, there exists a large and diverse personality assessment literature attesting to construct-validated connections between assessment measures and the real world that has been overlooked by D'Andrade and Shevitz (e.g., Block, 1957; Block, 1971, pp. 159-168, 218-229; Block, Jennings, Harvey, & Simpson, 1964; Cloninger, 1968; Olson, 1971; Jones, 1968; 1971; Katz, 1974; MacKinnon, 1962; Manheimer & Mellinger, 1967; Peckin, 1973; Robins, 1966; Rosenthal, 1970). A priori, it would appear difficult for a conceptual similarity viewpoint to explain the existence of relationships of the sort these, sampled from literally hundreds, that could have been offered.

The Need to Deepen, Clarify, and Substantiate the Meaning of Conceptual Similarity

The conceptual similarity position would replace the assumption that personality parameters (traits, dispositions) importantly underlie personality ratings with a linguistic explanation. Leaving aside the empirical relationships that the conceptual similarity interpretation must attempt to encompass, we note that this preferred explanation does not resolve the problem of understanding; it merely transfers or translates the problem. We must still face up to the thorny question of the nature and basis of similarity judgments. The study of similarity is a profoundly complex area of conceptual and empirical inquiry (Tversky, 1977); the invocation of "conceptual similarity" as an explanation without immediate close consideration and articulation of the processes by which similarity judgments are made is not especially helpful. D'Andrade was circumspect in his final claims, acknowledging that because of obscurities surrounding the way in which similarities are developed and recognized, it is somewhat ironic that the problem of the nature of similarity ratings and personality inventories on the basis of their conceptual similarity arguments and analyses, D'Andrade and Shevitz appear to have presumed the essential validity of personality assessment procedures.

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Shweder was less cautious in his view. He argued that judges make similarity judgments about the sort of individuals people are and how they may be used as measuring instruments. . . . As for the reasons why (personality or conceptual similarity) ratings provide information about people, the essay in the form of a brief exposition about exactly how these invalid results come about. (1974, p. 160)

Finally, we note that although the effects attributable to manipulation of the redundancy and structure of the variables being rated are large, the range of correlations between an individual's Conceptual Similarity matrix and the Rated Behavior matrix is interestingly wide. Moreover, the phenomenon previously noted (Block, 1971) was again observed: A judge's degree of consensus in conceptual similarity judging, even for the Approximately Orthogonal set, was related to his ability to "retrieve" the Approximately Orthogonal Rated Behavior matrix (r = .46, p < .05).

Separately, we will be studying the possible implications of these individual differences. There are preliminary indications that these differences are reliable and can provide additional perspective on the processes involved in the effort to develop, conceptualize, integrate, and apply one's psychological understanding of people. Meanwhile, we suggest that the inherent intricacy characterizing both human judgment and the ongoing stream of interpersonal behavioral makes it unlikely that so simple a principle as conceptual simi-
provide an adequate explanation complex a process as personality assessment

References

F. Interaction process analysis. Interna-encyclopedia of the Social Sciences, 1968,
Lives through time. Berkeley, Calif.: Ban-
1971. Advancing the science of personality: Para-shift or improving the quality of re-
M. Magnusson & N. S. Endler (Eds.) The psy-
ology at the crossroads: Current issues in onal psychology, Hillsdale, N.J.: Erlbaum, 1977. (a)
An illusory interpretation of the first fac-
MMP: A reply to Shweder. Journal of Consulting and Clinical Psychology, 1971, 45, 930-
H., Jennings, P. H., Harvey, E., & Simp-interaction between allergic potential and in- childhood asthma. Psychoso-
E. F., Cottrell, L. S., & Mann, J. H. The
the nature of individual interaction character-
M. Toward a cognitive social learning re-conceptualization of personality. Psychological Re-
1973, 80, 252-253.
Newcomb, T. M. The consistency of certain estro-

Man, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer-nomination personality ratings. Journal of Abnormal and Social Psychology, 1963, 66, 574-
Robins, L. Deviant children grown up. Baltimore, Md.: Williams & Wilkins, 1966.
Shweder, R. A. Illusory correlations and the MMI personality controversy. Journal of Consulting and Clinical Psychology, 1977, 45, 917-924. (a)
Shweder, R. A. Illusory correlation and the MMI personality controversy: Reply to some of the allusions and elisions in Block's and Edwards' commentaries. Journal of Consulting and Clinical Psychology, 1977, 45, 917-924. (b)

Postscript

1. If personality ratings demonstrate conceptually expected concurrent and predictive validity, the semantic "retrieval" of their intercorrela-
tions by Shweder and D'Andrade (1979) takes on a very different implication and significance. We cited a sampling of evidence, from a variety of sources, that ratings can generate replicable external relationships that support their empirical validity. Shweder and D'Andrade state that evidence of empirical validity is "beside the point" (p. 1079) in evaluating the question of the presence or absence of conceptual factors in judges we do not think so.

2. Interested readers will simply have to judge for themselves, by referring to the original sources, whether the behavior categories words differently from the rating dimensions "still main tain equivalent meaning." (p. 1078) as Shweder and D'Andrade declare.

3. Because we are without the space for proper psychometric explanation, we can only reiterate that the studies reported by Shweder and D'Andrade "provide no (or technically unsound reliability coefficients.)"

4. Shweder and D'Andrade remain unimpressed by the generality of their finding that appreciable intermatrix correspondence between what they called "actual behaviors" and "rated behaviors," was not observed. We suggest that the common effect of the many problems besetting their analysis is to increase unreliability and invalidity of measures.

5. In our study, we argued that retrieval of the matrix of correlations among Rated Behaviors by a matrix of relations among Conceptual Similarity judgments was a particular mix of personality variables employee namely, their bipolar psychological redundance. Shweder and D'Andrade (1979) suggest that reduction of the variance of the distribution the intervariable correlations creates an "artifact effect" (p. 1077). But, equally, the intermatrix correlations previously reported by Shweder and D'Andrade can be viewed as
"artifact effect" of extended variances of the distribution of intervariable correlations. Properly, arguments should not arise about artifactual effects on intermatrix congruences; the interdependent intervariable correlations from which they are computed have no "natural" population to which they can be referred so that one can sensibly talk of artificiality or representativeness of their distribution. What is important is that we have been able to provide a psychological explanation, in terms of redundancy and the proper measurement goals of investigators, of how these different extremes of "artificiality" come about and have their effect. In particular, high redundancy, both antonymous and synonymic, often has been deliberately introduced by investigators for good reasons (to balance direction of wording effects, to achieve psychometric respectability of subsequent composite scores, etc.). To make such redundancy subtle or indirect would be confusing to the psychologists using these rating scales. It is therefore not surprising that conceptual similarity judges are able to latch onto the structure underlying cleanly separated clusters of variables when highly and simply organized variable sets are employed. But such congruences are essentially adventitious; they have no implications, in and of themselves, for the validity and usefulness of well-based and well-encoded personality ratings.