

## Accurate Reflection or Systematic Distortion? A Reply to Block, Weiss, and Thorne

Richard A. Shweder  
Committee on Human Development  
University of Chicago

Roy G. D'Andrade  
Department of Anthropology  
University of California, San Diego

The "systematic distortion" hypothesis, advocated by D'Andrade and Shweder, states that correlational structures derived from memory-based personality ratings are primarily a product of the conceptual affiliations among rating categories rather than a reflection of the empirical correlational structure of behavior. This hypothesis is clarified with reference to the Block, Weiss, and Thorne critique. The "bipolar redundancy" effect, presented by Block, Weiss, and Thorne as an alternative explanation for correspondence between the correlations derived from memory-based ratings and judgments of conceptual affiliation, is found to be an artifact of reduced variance in matrices of uncorrelated traits. Block, Weiss, and Thorne argue that D'Andrade and Shweder failed to find correspondences between correlations for memory-based ratings and immediate scorings because of methodological flaws in their studies. This argument is reviewed, and it is concluded that the systematic distortion hypothesis remains well supported and that the evidence for the existence of covarying, multibehavior personality traits established by the correlational analysis of memory-based ratings remains dubious. Finally, an attempt is made to clarify the different uses of the term *trait* and to identify the different types of consistency in behavior that are relevant to the controversy.

Most classifications of individual differences in the personality literature are classifications of response patterns on personality rating forms, inventories, and questionnaire interviews. The pattern of correlations among variables in most of these classifications can be reproduced or replicated by asking a small number of respondents to judge the conceptual similarity or dissimilarity of the variables on the rating form, inventory, or questionnaire. To date, results from the following analyses of rating data have been successfully reproduced:

1. Factor-analytic classification of personality adjectives, as given in Norman (1963). (See D'Andrade, 1965; Mulaik, 1964.)

2. Leary grid organization of interpersonal behavior, as given in LaForge and Suzcek (1955). (See D'Andrade, 1965.)

3. Factor-analytic classification of personality and interpersonal behavior, as given in Bales (1970). (See Shweder, 1972, 1975.)

4. Factor-analytic classification of maternal personality, as given in Sears, Maccoby, and Levin (1957). (See Shweder, 1975.)

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

6. Partial correlation matrix for observers' ratings of extraversion-introversion in boys in summer camp, as given in Newcomb (1929). (See Shweder, 1975, 1977a.)

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

8. The common factor structure of the

---

Requests for reprints should be sent to Richard A. Shweder, Committee on Human Development, University of Chicago, 5730 South Woodlawn Avenue, Chicago, Illinois 60637, or Roy G. D'Andrade, Department of Anthropology, University of California at San Diego, La Jolla, California 92093.

Murray Needs from five personality tests, as given in Fiske (1973) and in Huba and Hamilton (1976). (See Ebbesen & Allen, Note 1.)

9. Syndrome clusters from the Brief Psychiatric Rating Scales, as given in Overall, Hollister, and Pichot (1967). (See Shweder & D'Andrade, Note 2.)

In these studies a variety of techniques for obtaining similarity judgments have been used, ranging from pile sorting to triads testing. The phrasing of the instructions for these judgments has also varied from an emphasis on "co-occurrence" to an emphasis on "likeness in meaning." Similarity judgments appear to be robust, in that small differences in technique do not seem to affect the correspondence found between conceptual similarity judgments and memory-based personality ratings.

At least two hypotheses can be constructed to account for the fact that conceptual similarity judgments reproduce the correlational structure of variables in personality ratings. The first is the "accurate reflection" hypothesis, which 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 hypothesis, people use empirically valid implicit personality theories in making conceptual similarity judgments, thereby accurately reporting the intercorrelation of behaviors (Jackson, Chan, & Stricker, in press; Passini & Norman, 1966).

The second hypothesis is the "systematic distortion" hypothesis, which asserts that the correspondence between preexisting ideas of what is like what and memory-based rating correlations occurs because, under difficult memory conditions, people infer what happened from their general model of what the world is like, and because conceptually affiliated memory items are easier to retrieve (Mandler, 1970). According to the systematic distortion hypothesis, the schemata held by most people tend to be inaccurate with respect to how behaviors covary, confusing "what is like what" with "what goes with what." Hence, memory for events contains a systematic bias, in that things that are con-

ceptually similar are recalled as if they covaried. L. J. Chapman (1967) has called this effect "illusory correlation."

A test of the accurate reflection hypothesis versus the systematic distortion hypothesis can be made by obtaining both observational evidence and memory-based ratings for a set of behavioral variables. By observational evidence we mean reasonably objective records made at the time of observation. Such records are called here "immediate scorings." (Block, Weiss, and Thorne, 1979, refer to this type of data as "Actual Behavior matrices.") To the extent that the correlations among variables derived from memory-based ratings are not like the correlations among variables derived from immediate scorings, but are like the pattern of conceptual similarity judgments, the accurate reflection hypothesis is disconfirmed, and the systematic distortion hypothesis supported. (See Shweder, 1977c, Note 3, for other ways to test the accurate reflection and systematic distortion hypotheses.)

In previous papers (D'Andrade, 1973, 1974; Shweder, 1975, 1977a) we examined published sets of data in which immediate scorings and memory-based ratings for the same variables were available. In all cases the pattern of intercorrelations for memory-based ratings was like the pattern of conceptual similarity judgments but only weakly related to the pattern of intercorrelations for the immediate scorings. The evidence supported the systematic distortion hypothesis, and we concluded that personality classifications derived from memory-based data could not be trusted.

Block, Weiss, and Thorne (1979) assert that the systematic distortion hypothesis is not adequately supported. Two general arguments are advanced: the "bipolar redundancy" argument and a series of methodological criticisms.

#### *The Bipolar Redundancy Argument*

Block, Weiss, and Thorne (1979) claim that the correspondence between conceptual similarity matrices and correlation matrices for memory-based ratings is "a previously unrecognized function of the high redundancy and

the structure of that redundancy that just happened to be present in the particular variable sets D'Andrade and Shweder sampled" (pp. 1065-1066). They attempt to show this by demonstrating that by varying the degree of bipolar redundancy among variables, they can affect the degree of correspondence between conceptual similarity matrices and correlation matrices for memory-based behavior ratings. Using the California Q-set variables and a factor analysis based on ratings by clinical psychologists of a sample of 160 subjects, they selected three sets of 12 items each. For the first set, 6 items with high positive loadings and 6 items with high negative loadings from a single factor were selected; for the second set, 2 items with high positive loadings and 2 items with high negative loadings from three orthogonal factors were selected; for the third set, single items with high loadings from each of 12 approximately orthogonal factors were selected.

For each of these three sets, conceptual similarity judgments were made by a small sample of respondents. For the first set, the mean correlation between the individual conceptual similarity matrices and the rated behavior matrix was found to be .63. For the second set, the mean correlation was found to be .50. But for the third set, which is made up of items from orthogonal factors, the correlation was found to be only .21.

Based on the fact that the mean intermatrix correlation for the orthogonal items is significantly less than the mean intermatrix correlations for items selected from bipolar factors, Block, Weiss, and Thorne (1979) conclude that "Conceptual Similarity - Rated Behavior intermatrix correspondence is high when bipolar redundancy is high and is low when bipolar redundancy is low" (p. 1068) and that they have therefore controlled the degree of correspondence between conceptual similarity and memory-based rating correlations by varying the degree of bipolar redundancy.

We believe that Block, Weiss, and Thorne are incorrect in this claim, since the artifactual effect of the reduced variance within the matrix of orthogonal items has been overlooked. Using Block, Weiss, and Thorne's original data, kindly supplied to us by the authors,

and comparing the three data sets, we found standard deviations of .53 and .37 in the rated behavior correlation matrices for the first and second sets, in contrast to a standard deviation of only .15 for the third set. The variability of the rated behavior correlations in the third set has been much reduced by the selection process, and this in turn has reduced the possible covariance between the matrices.

It is also crucial to note that contrary to Block, Weiss, and Thorne's claim, *respondent judgments about the conceptual similarity of the items in the third set do in fact correspond to the level of the correlations in the rated behavior matrix*. Our reanalysis of Block, Weiss, and Thorne's data reveals that the mean conceptual similarity score for items in the third set is 4.3 on a scale for which 1 is defined as "extremely dissimilar," 4 as "unrelated," and 7 as "extremely similar." The mean of the correlations in the rated behavior matrix is .12 disregarding sign. Thus, in a set of almost orthogonal items, conceptual similarity judgments indicate that the items are almost unrelated.

The problem with using correlations to measure the relation between matrices in this instance can be illustrated by considering the extreme case in which all the items are correlated .00 with each other, and all the conceptual similarity judgments are at 4.0, indicating that the items are unrelated. Using correlations as the only measure of correspondence, one would assume that there was no relation between these matrices, even though the judgments and the correlations match perfectly. We conclude that the correspondence between conceptual similarity matrices and correlation matrices for memory-based ratings is not a function of adventitious redundancy.

There is an important methodological problem with the redundancy hypothesis, namely, that the independent variable is operationally the same as one of the dependent variables. By "redundancy" Block, Weiss, and Thorne mean the "communal variance" shared by a set of variables. But communal variance is a function of the intercorrelations between the variables. Therefore, in trying to control for redundancy, they are in effect controlling for

the intercorrelations between variables, which is also one part of the relation they are trying to predict. As a result, in reducing redundancy they necessarily reduce the variance in the intercorrelations between the ratings and, thereby, the possible covariance between matrices.

It is interesting that the redundancy hypothesis asserts part of what the systematic distortion hypothesis asserts: that there is a correspondence between conceptual similarity judgments and the intercorrelations among memory-based behavior ratings (since redundancy is a function of the intercorrelation among variables); however, it reverses the direction of causation and thus becomes a version of the accurate reflection hypothesis.

### *The Methodological Critique*

Block, Weiss, and Thorne (1979) present a series of methodological criticisms concerning the analyses by D'Andrade (1974) and Shweder (1975) of data taken from studies by Newcomb (1929), Borgatta et al. (1958), and Mann (1959). What follows is a list of responses to these criticisms.

1. "The categories or items or dimensions of behavior measured by the Actual Behavior indices are often different from those rated by observers" (Block, Weiss, & Thorne, 1979, p. 1060). This is an important point, since the data were analyzed as if they were from a memory experiment in which the task is to try to remember the relative frequencies of various behaviors across different actors. To the extent that the memory-based ratings were made on categories not equivalent in meaning to the categories used in the immediate scorings, lack of correspondence between the two sets of data becomes artifactual.

Since our analyses were carried out on already-published data, we had no control over how the labeling was done. In our judgment, the categories that do not have identical wording still maintain equivalent meaning. However, it is possible to make a more stringent test by using only categories that have identical wording. We have recomputed the intermatrix correlations for the Mann data using only items that are identically phrased (Categories 2, 3, 4, and 7). The intermatrix corre-

lations between the immediately scored behavior matrices and memory-based rating matrices for the Mann data are .15,  $-.03$ ,  $-.27$ , and  $.03$ , respectively ( $M = -.03$ ), with the nonidentical items removed, compared to intermatrix correlations of  $.07$ ,  $.20$ ,  $-.03$ , and  $.27$ , respectively ( $M = .13$ ), for the data as originally given. Thus removing the items that were not identically phrased did not increase (in fact decreased slightly) the correspondence between the actual behavior and the memory-based rated behavior matrices.

Since the issue about maintaining equivalent definitions in both conditions is important for evaluating the evidence for or against the systematic distortion hypothesis, it should be mentioned that relevant data exist in an unpublished study (Shweder & D'Andrade, Note 2). Block, Weiss, and Thorne did not have the results of this study available to them at the time they wrote their critique, nor do the results of this new study diminish the pertinence of their concern about this issue. Briefly, this study is similar in design to the studies already reviewed, except that (a) the categories of behavior were taken from natural language terms for social interaction (e.g., "advises," "criticizes," etc.), (b) the immediate scoring categories and the memory-based rating categories are identical, and (c) the stimulus material consisted of videotape of natural family interaction, thereby permitting immediate scoring by a number of observers. The intermatrix correlations are quite similar to the results found in the other cases: The correlation between the immediately scored behavior matrix and the memory-based behavior rating matrix is  $.23$ ; the correlation between the memory-based behavior rating matrix and the conceptual similarity matrix is  $.75$ ; and the correlation between the immediately scored behavior matrix and the conceptual similarity matrix is  $.00$ .

In summary, on the basis of the recomputed figures and the new study, we conclude that the low degree of correspondence we have consistently found between memory-based rated behavior correlation matrices and immediately scored behavior correlation matrices is not a product of wording changes in behavioral categories.

2. "Emphasis solely on the correlations between traits ignores consideration of the validity of specific traits" (Block, Weiss, & Thorne, 1979, p. 1061). This is true, but beside the point. The reason for the emphasis on the correlations between traits is that it is these correlations that test the systematic distortion and accurate reflection hypotheses. Of course, if the correspondence between the memory-based data and the immediately scored data were very high, there would not be enough error variance for systematic distortion to take place. However, the correlations between specific immediate scorings and memory-based behavior ratings are in the .3 range. Even vigorous correction for attenuation would not put the level of validity so high that there would be insufficient error variance to permit the effect of systematic distortion to occur.

Related to the issue of validity, Block, Weiss, and Thorne (1979) raise questions about the reliability of the immediate scorings. They assert that "because the three studies provide no (or technically unsound) reliability coefficients, we cannot estimate what the relationships between the Rated Behavior and the Actual Behavior variables would be if they were psychometrically improved" (p. 1062). For the Newcomb (1929) study, an odd-even day reliability check on behavior percentages yielded a mean reliability coefficient of .78 over 26 variables. In the Borgatta et al. (1958) study, reliability was based on comparing the profiles of the three groups across two sessions, with the results presented graphically. From visual inspection there appears to be high consistency in the results for each group across sessions. In the Mann (1959) study, Mann did the scoring alone. However, he had worked with R. F. Bales for 3 years and had attained a scoring reliability of approximately .90 on the majority of the categories, with Bales as the criterion. However, since the reliabilities are not given on a category-by-category basis, Block, Weiss, and Thorne are correct in pointing out that correction for attenuation cannot be made. Nevertheless, the figures indicate that the immediate scorings have reasonable mean reliability in all three studies.

3. "The use of frequency counts of behav-

ior is not a sufficient means of operationalizing complex psychological concepts" (Block, Weiss, & Thorne, 1979, p. 1062). Again, this is true but beside the point. This criticism would be appropriate were we trying to develop a personality assessment instrument. Our main concern in these studies is not to develop personality assessment techniques but to test hypotheses about systematic distortion in memory. In trying to anchor one set of ratings in what can be observed and counted, we have used studies that used relatively simple and direct methods. If we did not have simple and direct measures, how could we know whether memory distortion had occurred?

This same general point holds for the related criticism by Block, Weiss, and Thorne that the categories of Interaction Process Analysis (IPA) are not the most effective means of describing social interaction and (on the other side of the fence) the criticism that the Newcomb categories require considerable psychological inference. Difficulties in applying particular category systems would be likely to lower the validity of the ratings, but would not introduce systematic distortion.

Block, Weiss, and Thorne (1979) also question the use of averages as a way of summarizing behavior. They say, "Useful though the averaging 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" (p. 1064). We do not deny these criticisms of averages, but we do deny the relevance of these criticisms to our research. Repeating the point made above, such criticisms would apply if we were attempting to assess the personality of the subjects rated in these studies. Given that we are trying to test hypotheses about memory, we hold that these criticisms do not apply, since the methods used did not artifactually introduce systematic distortion. Were it the case that the memory-based raters captured important information not to be found in the immediate scorings, this would affect the correlations between specific categories across conditions but would not affect the relation between the correlation matrices.

4. "Extraneous influences distorting correlation coefficients and their patterning are not considered" (Block, Weiss, & Thorne, 1979, p. 1064). First, Block, Weiss, and Thorne make the point that the IPA rates are likely to be highly skewed or not unimodal, which will tend to lower correlations. Second, where the behavior counts are transformed into percentages, some degree of negativity is introduced into the correlation matrix. Third, in four of the five sets of data, the memory-based behavior rating data were based on pooled judgments from a number of observers, but the immediately scored behavior data were obtained from a single observer, so that the discrepancies may be a function of the "volatility" of the immediate scorings.

If we had merely presented one test of the systematic distortion hypothesis that had all of these problems, our results would be equivocal. However, we have presented a number of different tests of the systematic distortion hypothesis on different types of data, collected by different investigators, using different methods. One advantage of using a number of different tests is that criticisms of the data or of the procedures in one case can be matched with a case in which the criticism does not apply. We can then see if the same results obtain. Thus the Newcomb (1929) data do not have distributional difficulties of the sort that might have affected the Mann (1959) data, and these data are also not affected by the negativity problem. Further, in the Mann study there are two sets of data—one in which there is no discrepancy between the number of subjects used in the memory-based behavior rating condition and the number of subjects used in the immediately scored behavior condition, and one in which there is. No difference is observed in the results. Thus, to each of Block, Weiss, and Thorne's three criticisms there are one or more tests to which the criticism does not apply, yet the same pattern of findings occurs. Almost every study has some methodological weak point. We believe it is the consistency with which a given result is obtained across variations in method that ultimately supports or disconfirms a hypothesis.

Also included under the heading of "ex-

traneous influences" by Block, Weiss, and Thorne are two other criticisms. The first concerns the use of correlations to measure the degree of relatedness between two matrices of correlations. Block, Weiss, and Thorne point out that because of the lack of independence in the "observations," the normal interpretations of the correlation coefficient with regard to amount of variance explained and levels of significance are inappropriate. To date we have not found a better way to compare these types of correlation matrices. However, even under these conditions, the correlation coefficient maintains its descriptive value, yielding an interpretable figure, which indicates how well one can predict from one to the other set of measurements. Since we have no alternative, and since we are primarily interested in obtaining a comparative figure for the *relative* degree of association between various sets of measurements, we have perforce relied on the product-moment and Spearman's rank-order correlation coefficient.

The second issue concerns the effect of reflecting variables in a matrix of correlations, thereby changing the degree of association between matrices. Since some psychological variables permit such reflections, it is possible that some of the variables could have been reflected, which would possibly have changed the relationship between matrices. However, this criticism applies only to some of the studies that attempt to reproduce the structures found in various personality assessment instruments. None of the variables we have used in comparing immediately scored behavior matrices with memory-based behavior rating matrices were of the dimensional or bipolar types that would permit reflection, and so this particular problem does not affect our attempts to test the systematic distortion hypothesis.

Last on the list of methodological criticisms, Block, Weiss, and Thorne (1979) raise the question of whether the correlations found between the immediately scored behavior matrices and memory-based behavior rating matrices are so high that it is impossible to "unconfound the conceptual similarity interpretation from the isomorphism view" (p. 1065). They say, "Instead of being impressed that

these 'coefficients' are so low, one can as well be surprised that they are so high" (p. 1065). However, the test of the systematic distortion versus accurate reflection hypothesis requires the comparison of *three* correlations: (a) the correlation between the memory-based behavior rating matrix and the conceptual similarity matrix, (b) the correlation between the memory-based behavior rating matrix and the immediately scored behavior matrix, and (c) the correlation between the immediately scored behavior matrix and the conceptual similarity matrix.

Across seven different tests we find the following: (a) The mean correlation between memory-based behavior rating matrices and immediately scored behavior matrices is .25; (b) the mean correlation between memory-based behavior rating matrices and conceptual similarity matrices is .75; and (c) the mean correlation between the immediately scored behavior matrices and conceptual similarity matrices is .26.

Comparing the three relationships, we observe that the relation between the matrices for memory-based ratings and the matrices for immediate scorings is weaker than the relation between the matrices for memory-based ratings and the conceptual similarity matrices. This result is not due to some effect of the relation between the immediate scoring matrices and the conceptual similarity matrices, since these are also only weakly related. It is this pattern of correlations that supports the systematic distortion hypothesis, not our impression of whether any one of these relationships is surprisingly strong or weak.

### *Theoretical Issues*

In the final section of their article, Block, Weiss, and Thorne take up a series of general theoretical issues concerning the conceptual similarity position. One of the issues they raise concerns the nature and the 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" (Block, Weiss, & Thorne, 1979, p. 1071). The study of similarity judgments has been a neglected area. A simple model, informally presented, dealing with only the denotative aspects of meaning, is given in Romney and D'Andrade (1964). It is our current view that similarity judgments are based on the relative number of predications common to the two terms being rated; such predications we believe refer to a wide variety of things, including the sets to which both items belong, parts which are common to both items, common effects of both items, and so on. The study of similarity can be divided into two parts: The first is to determine how information is processed in making a judgment; the second is to determine what informational content is being processed. Multi-dimensional scaling of similarity judgments has proven to be an effective way of investigating the second problem (Romney & D'Andrade, 1964; Shepard, 1974). It seems likely that with the stimulus of Tversky's model there will soon be an increase in research on the first problem.

Another general issue taken up by Block, Weiss, and Thorne (1979) is the degree to which cultural belief systems can be assumed to be veridical. They point out that although constructs need not be empirically accurate, they often are. Further, they argue that there is good reason to expect cultures and individuals to develop both reliable and valid ways of describing human personality (p. 1071).

As anthropologists interested in the relation between culture and cognition, we support the generalization that cultural belief systems are sometimes empirically accurate. However, having some experience with the strong commitment people often have to beliefs that are empirically inaccurate, we feel that the other side of the proposition also needs stating; that is, cultural belief systems can be quite inaccurate and yet quite compelling. In this instance, the question is whether people have personality traits consisting of covarying behaviors. The fact that this is believed to be true in our culture does not constitute evidence for either its truth or its falsity.

From a more psychologically oriented per-

spective, Block, Weiss, and Thorne (1979) assert: "In seeking to dispense with personality ratings and personality inventories on the basis of their conceptual similarity arguments and analyses, D'Andrade and Shweder appear to have presumed the essential invalidity of personality assessment procedures" (p. 1070). Arguing for the essential validity of personality assessment procedures, Block, Weiss, and Thorne introduce "some relationships of a kind that personality psychologists find reinforcing of their belief in reliable individual differences with regard to personality parameters of consequence in the real world" (p. 1069). As examples of relationships of this kind they point to across-time consistencies in trait ratings and the ability of personality assessment measures to discriminate between criterion groups. Such relationships, they assert, can be neither reproduced nor explained by the conceptual similarity position.

It is true that such relationships, to the extent they did exist, could not be reproduced or explained by reference to conceptual similarity. This is because the conceptual similarity argument has no bearing on the issue of perceived longitudinal consistency or on the asserted ability of personality measures to discriminate between criterion groups.

What becomes apparent in examining the final section of Block, Weiss, and Thorne's article is that there are (at least) two uses of the term *trait* and the related phrase *individual differences*. The conceptual similarity position is part of an attack on the hypothesis that people have global traits and on the related hypothesis that one can determine such traits through the analysis of individual differences. The defense given by Block, Weiss, and Thorne concerns the predictive utility of traits—but in a different sense of the word—and the possibility of determining such traits through the analysis of individual differences—in a different sense of the phrase.

Originally, the notion of a trait appears to have contained four distinct kinds of consistency: consisting across similar behaviors (an honest person does not lie, does not steal, and does not cheat); consistency across situations (an honest person does not cheat when taking a math test and does not cheat when playing

a ring toss game); consistency across time (an honest person is likely to have been honest yesterday and will be honest tomorrow); and consistency within a framework of psychological functioning (honesty is theoretically and empirically related to self-control and identification with parental figures). (See McClelland, 1951, 210–234.)

The kind of trait consistency that is relevant to the conceptual similarity position is consistency across similar behaviors. The classical approach to the problem of determining consistency across similar behaviors is correlational analysis, in which the correlations are computed across persons, as typified by the work of Eysenck and Cattell. This approach assumes that individual differences arise from differences between persons in the amount of a trait, and so the co-variation of behaviors across persons can be used to determine which behaviors belong to which traits (Cattell, 1946).

It has been our position that evidence for the existence of this type of trait—this type of consistency across similar behaviors—is rather dubious. Most of the supporting data are from memory-based ratings, questionnaires, and inventories. (Correlational data from self-report tests can be considered "objective evidence" in one sense, but such data do not establish that the behaviors reported on actually co-occur as rated.) We believe that memory-based data are so strongly affected by conceptually based systematic distortion that they are primarily evidence for the existence of a powerful cultural belief system.

However, the definition of *trait* given above does not correspond to the use of the term by Block, Weiss, and Thorne (1979). They say:

Personality psychologists seek or settle on useful sets of variables to be employed for the specification of the personal or behavioral qualities of the individuals to be studied. To the extent that the set of personality or behavioral variables has communal variance (i.e., is redundant), there will be a "similarity structure of the trait measurements." But such similarity structures represent neither conceptual nor empirical fixities. (p. 1056–1057)

Thus, for Block, Weiss, and Thorne, the emphasis is on useful ratings, which usually involve constructing a composite index across

a number of measures. The different measures to be placed in a composite index do not have to be internally correlated or homogeneous; they may even lack face validity, as long as the composite index displays significant relationships with an external criterion.

With respect to current personality assessment work as described by Block, Weiss, and Thorne, it appears that two of the four types of consistency originally contained in the notion of a trait have been dropped or suspended: consistency across similar behaviors and consistency across different situations. If true, this is an interesting shift. We have been arguing that one of these two suspended types of consistency has little empirical warrant, and we are encouraged to believe that this position no longer appears heretical.

#### Reference Notes

1. 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, California 92037.)
2. Shweder, R. A., & D'Andrade, R. G. *Personality or culture?* Unpublished manuscript, 1978. (Available from R. A. Shweder, Committee on Human Development, University of Chicago, Chicago, Illinois 60637, or R. G. D'Andrade, Department of Anthropology, University of California at San Diego, La Jolla, California 92093.)
3. Shweder, R. A. *Fact and artifact in personality assessment: The influence of conceptual schemata on individual difference judgments.* Paper presented at the 86th annual convention of the American Psychological Association, Toronto, Canada, August 1978. (Available from R. A. Shweder, Committee on Human Development, University of Chicago, Chicago, Illinois 60637.)

#### References

- Bales, R. F. *Personality and interpersonal behavior.* New York: Holt, Rinehart & Winston, 1970.
- Block, J. *The challenge of response sets.* New York: Appleton-Century-Crofts, 1965.
- Block, J., Weiss, D. S., & Thorne, A. How relevant is a semantic similarity interpretation of personality ratings? *Journal of Personality and Social Psychology*, 1979, 37, 1055-1074.
- Borgatta, E. F., Cottrell, L. S., & Mann, J. H. The spectrum of individual interaction characteristics: An interdimensional analysis. *Psychological Reports*, 1958, 4, 279-319.
- Cattell, R. D. *Description and measurement of personality.* New York: World Book, 1946.
- Chapman, L. J. Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 151-155.
- D'Andrade, R. G. Trait psychology and componential analysis. *American Anthropologist*, 1965, 67, 215-228.
- D'Andrade, R. G. Cultural constructions of reality. In L. Nader & T. W. Maretzki (Eds.), *Cultural illness and health.* Washington, D.C.: American Anthropological Association, 1973.
- D'Andrade, R. G. Memory and the assessment of behavior. In T. Blalock (Ed.), *Measurement in the social sciences.* Chicago: Aldine-Atherton, 1974.
- Fiske, D. W. Can a personality construct be validated empirically? *Psychological Bulletin*, 1973, 80, 89-92.
- Huba, G. J., & Hamilton, D. L. On the generality of trait relationships: Some analyses based on Fiske's paper. *Psychological Bulletin*, 1976, 83, 868-875.
- Jackson, D. N., Chan, D. W., & Stricker, L. J. Implicit personality theory: Is it illusory? *Journal of Personality*, in press.
- LaForge, R., & Sucek, R. F. The interpersonal dimension of personality: III. An interpersonal checklist. *Journal of Personality*, 1955, 24, 94-112.
- Mandler, G. Words, lists and categories: An experimental view of organized memory references. In J. L. Cowan (Ed.), *Studies in thought and language.* Tucson: University of Arizona Press, 1970.
- Mann, R. D. *The relation between personality characteristics and individual performance in small groups.* Unpublished doctoral dissertation, University of Michigan, 1959.
- McClelland, D. C. *Personality.* New York: Sloane Associates, 1951.
- Mulaik, S. A. Are personality factors raters' conceptual factors? *Journal of Consulting Psychology*, 1964, 28, 506-511.
- Newcomb, T. M. The consistency of certain extravert-introvert behavior patterns in 51 problem boys. Contributions to Education No. 382. New York: Teachers College, Columbia University, 1929.
- Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 1963, 67, 574-583.
- Overall, J. E., Hollister, L. E., & Pichot, P. Major psychiatric disorders: A four-dimensional model. *Archives of General Psychiatry*, 1967, 16, 146-151.
- Passini, F. T., & Norman, W. T. A universal conception of personality structure? *Journal of Personality and Social Psychology*, 1966, 4, 44-49.
- Romney, A. K., & D'Andrade, R. G. Cognitive aspects of English kin terms. *American Anthropologist*, 1964, 66, 146-170.
- Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing.* New York: Harper & Row, 1957.
- Shepard, R. N. Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 1974, 39, 373-421.

- Shweder, R. A. Semantic structures and personality assessment (Doctoral dissertation, Harvard University, 1972). *Dissertation Abstracts International*, 1972, 33, 2452B. (University Microfilms No. 72-29,584)
- Shweder, R. A. How relevant is an individual difference theory of personality? *Journal of Personality*, 1975, 43, 455-484.
- Shweder, R. A. Illusory correlation and the MMPI controversy. *Journal of Consulting and Clinical Psychology*, 1977, 45, 917-924. (a)
- Shweder, R. A. Illusory correlation and the MMPI controversy: Reply to some of the allusions and elusions in Block's and Edwards' commentaries. *Journal of Consulting and Clinical Psychology*, 1977, 45, 936-940. (b)
- Shweder, R. A. Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. *Current Anthropology*, 1977, 18, 637-648. (c)
- Tversky, A. Features of similarity. *Psychological Review*, 1977, 84, 327-353.

Received December 24, 1978 ■