

Personality Traits: Fact or Fiction? A Critique of the Shweder and D'Andrade Systematic Distortion Hypothesis

Daniel Romer
University of Illinois at Chicago

William Revelle
Northwestern University

According to Shweder and D'Andrade (1979, 1980), covariation in memory-based ratings of people's behavior is determined more by semantic relations between behavior categories than by actual co-occurrence. They claim therefore that the existence of personality traits is largely a fiction supported by our conceptions rather than by reality. Contrary to this hypothesis, we argue that semantics are logically implicated in both the observation and recall of behavior and that support for this assumption can be found if immediate encodings of behavior are as sensitively scaled as subsequent memory-based ratings. Results of a demonstration experiment supported this conclusion. When immediate encodings were scaled across all behavior categories, the relation between semantics and memory was completely explained by the role of semantics in the immediate encoding of behavior. However, when immediately encoded behavior was simply identified (rather than scaled), support for systematic distortion was obtained. Previous support for the systematic distortion hypothesis may therefore be attributed to the use of too simple a coding scheme for the measurement of immediate behavior. Implications for the existence of personality traits and for personality measurement are discussed.

Whether personality traits truly exist or are merely artifacts in the minds of observers has long been a controversial question (e.g., Fiske, 1978; Mischel, 1968, 1973; Newcomb, 1931; Thorndike, 1920). Of the many uses of the trait concept, perhaps the most critical is the assumption that the numerous manifestations of a person's behavior can be subsumed by underlying stabilities in character or personality. Thus the many ways of expressing aggression might be traced to the existence of a stable characteristic or trait such as aggressiveness. Without this assump-

tion, the trait concept loses much of its scientific appeal. Writers such as Mischel have thrown considerable doubt on the assumption of stable traits by appealing to the situation specificity and plasticity of behavioral manifestations of traits. Others (e.g., Bowers, 1973; Endler & Magnusson, 1976) have emphasized the importance of the interaction between person variables such as traits and situations. Although the issue is far from settled, these attacks have placed the validity of the trait concept somewhat in question.

One very strong case against the trait concept can be derived from research on the attributional behavior of the layperson. A widely observed phenomenon, the fundamental attribution error (Jones, 1979), suggests that the layperson overestimates the contribution of personality characteristics as causes of behavior and similarly ignores the effects of situations. Thus it can be argued that the scientific appeal of traits actually reflects the layperson's tendency to overattribute underlying stability in others' behavior when actually behavior might largely be de-

This research was supported partly by Grant No. MH29209 from the National Institute of Mental Health (William Revelle, principal investigator). We thank T. Lederer for her assistance in preparing this manuscript, and L. Hazlewood, A. Jackson, R. McLaughlin, J. Onken, D. Parella, B. Park, K. Rasinski, and S. Siegel for participating as observers. Helpful comments from L. Alloy, J. Crocker, and R. Shweder on an earlier version of this article are gratefully acknowledged.

Requests for reprints should be sent to either Daniel Romer, Department of Psychology, Box 4348, University of Illinois at Chicago, Chicago, Illinois 60680, or William Revelle, Department of Psychology, Northwestern University, Evanston, Illinois 60521.

Table 1

Correlations (rs) Between Immediately Observed Behavior and Between Memory for Behavior, and Semantic Similarities (ss) Between Behavior Categories

Behavior			Memory				Semantics				
Category	A	B	C	Category	a	b	c	Category	α	β	γ
A	—			a	—			α	—		
B	r_{AB}	—		b	r_{ab}	—		β	$s_{\alpha\beta}$	—	
C	r_{AC}	r_{BC}	—	c	r_{ac}	r_{bc}	—	γ	$s_{\alpha\gamma}$	$s_{\beta\gamma}$	—

terminated by situations. This error might then be compounded by tendencies to make error-free predictions of future or related behavior from small samples of behavior and to draw these inferences on the basis of heuristics such as representativeness rather than careful attention to actual behavior rates (Kahneman & Tversky, 1973).

Shweder (1975, 1977) and D'Andrade (1965, 1974) advanced one extremely provocative interpretation of these attribution tendencies. According to their "systematic distortion" position, observers impose semantic structure on behavior, when in fact no such structure exists. That is, rather than recognizing the true empirical relations between behavior categories, observers use a representativeness or similarity heuristic to describe the relations between categories.

In support of their view that traits exist only in the minds of observers, Shweder (1975, 1977) and D'Andrade (1965, 1974) referred to the relations that obtain between immediately observed behavior, subsequent ratings of the same behavior and the semantic structure of the ratings. When frequency counts of behavioral observations (or encodings) are calculated for a set of individuals and those encodings are subsequently rated from memory by the same observers, it is possible to determine correlations between behavior categories based on either the behavior encodings or on the ratings. Furthermore, one can determine semantic relations between the various behavior categories by asking observers to report the perceived similarities between the categories. As shown in Table 1, these matrices have corresponding elements for their respective rows and columns, permitting the calculation of a similarity or correlation coefficient between matrices.

Shweder and D'Andrade (1979) reviewed the results of seven studies in which the relevant comparisons were made. These studies reveal that the behavior correlations are small and often near zero, whereas the memory-based (ratings) correlations are larger. Furthermore, the rating correlations appear to parallel the semantic similarities (average $r = .75$) to a greater degree than the behavior correlations (average $r = .26$). Finally, the immediately encoded behavior matrices correlate only about .25 with the respective memory-based matrices.

From data such as these, Shweder and D'Andrade (1979) concluded that (a) immediately observed behavior counts reveal little cohesion in the actual co-occurrence of behavior; (b) trait ratings based on recollection of observations contain considerable cohesion and structure that is closely related to the semantic connections between traits; and (c) observers relying on memory-based data commit the various attribution errors outlined earlier and impose a semantic similarity structure on the co-occurrence of behavior in the observed sample. Shweder and D'Andrade noted that these conclusions do not rule out the existence of stable patterns of individual differences; rather, they suggested that the belief that these patterns are organized into stable traits is not only invalid but is concocted by observers who are too overburdened by the data that must be remembered to render an accurate account of the true data pattern.

Shweder and D'Andrade (1980) did not leave the matter here, however, because it was not only memory burdens that were held responsible for inaccurate judgment of behavior co-occurrence. As anthropologists they were interested in the parallels between these conclusions and variation in cultural belief

systems that are sometimes empirically invalid. Just as some primitive cultures engage in what is known as *magical thinking*, the average civilized person "confuses propositions about likeness with propositions about co-occurrence likelihood" (Shweder, 1977, p. 642).¹ We attribute causes (e.g., traits) to events (e.g., behavior) based on their semantic similarity rather than on their empirical correlation. Thus "the failure to report accurately upon correlational relationships in one's experience is an indication of the absence of a concept of correlation in normal adults" (Shweder, 1977, p. 642). We are not simply poor tabulators of experience; we are ignorant of the basic concepts needed to draw the appropriate inferences from experience (see Crocker, 1981, for a review).

These conclusions have not gone unchallenged. Block, Weiss, and Thorne (1979) noted some difficulties in drawing these conclusions from Shweder and D'Andrade's (1979) data. For example, many factors such as the definitions of behavior can affect the correlations between behavior types resulting in a poor match between matrices. In this article we focus on some interpretive difficulties that we believe are sufficiently serious to throw great doubt on Shweder and D'Andrade's conclusions. In order to do so, we begin by setting out a model of the role that semantics play in the behavior-encoding process and in subsequent memory-based judgment.

Model of Behavior Encoding

Figure 1 outlines a structural model for the relations between behavior encoding, memory for behavior, and semantics. Structural modeling seems to be appropriate for discussing the relations between these processes inasmuch as the data Shweder and D'Andrade (1979) discussed are essentially correlations between measures of these processes. Furthermore, the use of structural modeling helps to clarify both the theoretical and measurement assumptions that enter into the conclusions that can be drawn from correlational data.

The starting point in the model is the covariation in behavior exhibited by a set of individuals. The assumption here is that be-

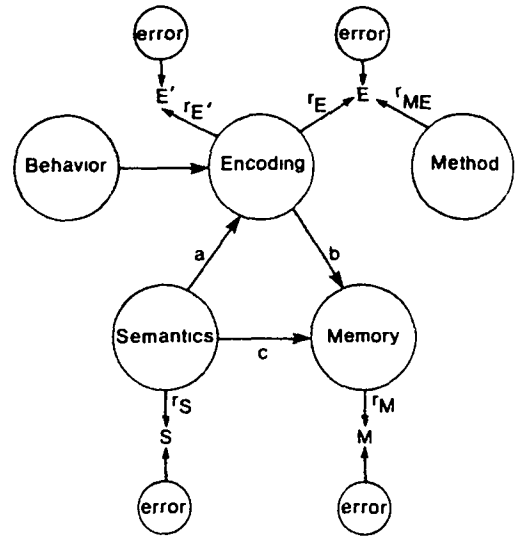


Figure 1 Structural model showing the underlying relationships between behavior, encoding, semantics and memory. (The observed measures of these latent variables are assumed to reflect both true and error variance. Latent variables are represented by circles, and causal relationships by arrows.)

havior is simply a set of physical movements or states; no label is attached to behavior until it is encoded by an observer. This assumption is a virtual truism in the social sciences and Shweder (1977) apparently accepted it. One implication of this assumption is that covariation in behavior is by definition unobserved and so is treated exclusively as a latent variable. Its first manifestation occurs at the encoding stage. This encoding is assumed to be a joint function of both behavior and the semantic structure that observers use in encoding it (Path A). To encode behavior such as "giving praise," for example, requires mapping rules between events or states (e.g., smiling, making positive remarks, patting someone on the back) and a behavior label. This is the province of semantics.

Semantics also define relations between behavior labels. Thus "giving praise" is perhaps similar to "offering encouragement" but

¹ Shweder (1977) gives examples of magical thinking in the therapeutic practices of the Azande who attempt to cure epilepsy by the administration of monkey skulls or to cure ringworm by the application of fowl excrement. These "cures" are said to be conceived because of their resemblance in some respect to the ailment.

different from "giving criticism." Indeed, one might say that the behavior-encoding process implicitly rests on such similarity judgments because the observer must decide how congruent a given behavior is with all of the many labels that might be matched to it before he or she arrives at the appropriate description. Although we accomplish this judgment with great ease, it is by no means understood how we do so, and the present model contains no assumptions about this process other than that it occurs. However, an implication of the model is that different semantic structures produce different encodings. Thus although the model is consistent with Shweder and D'Andrade's (1979) assumption that semantics determine how we describe reality, there is nothing magical about this process. Indeed, the effect of semantics is purely definitional in the sense that except for error in perceiving behavior,² semantics *cannot* be wrong at this stage.

The memory of this encoding process is presumably the impression that remains of the individuals observed (Path B). It is at this stage that trait relations are said to be constructed more by semantics than by true encodings (i.e., Path C is large whereas B is small). In addition, the apparently small size of B is said to be evidence for the conclusions that we are insensitive to the relations between behavioral events and that we have difficulty extracting the true correlational structure of our experience.

To this point we have discussed the implications of the systematic distortion position in terms of theoretical latent variables. Also included in the model, however, are the corresponding measures of encoding (E), memory (M), and semantics (S). These measures are imperfect in the sense that each of the latent variables is indexed by instruments that are subject to error. This error has at least two components: random noise and stable error. The former component is typically thought of in terms of the reliability of the measuring instrument's readings and is indexed by test-retest or interitem correlations. Arrows between the latent variables and their measures (e.g., r_S) represent this source of reliability.

The second component, however, reflects the inability of an instrument to capture the

true variable even if its reliability is perfect. In this case, the instrument's readings are partly determined by a stable error that is uncorrelated with the true latent variable of interest. The result is that there is variation in the true variable that cannot be detected by the instrument. This sort of error is usually conceived of as invalidity because the instrument will not enable one to predict a criterion as well as a more valid instrument will. In the absence of a more valid instrument, however, it is difficult to determine how valid an instrument is, and the error can be misattributed to a weak relation between the latent variable and a criterion variable. Because we hope to show that this source of error is responsible for the weak relation between encoding and the other variables, we include a case of a method factor in the measurement of encoding. Even if the reliability of E is large, its relation to the true encoding variable (r_E) will be attenuated by method error (r_{ME}). To the degree r_E is small, the relation between E and M or E and S will also be attenuated. However, a measure that is not subject to a method bias (E') should be a more valid indicator of encoding and hence should display higher correlations with the other variables in the model.

Shweder and D'Andrade (1979), on the other hand, assumed that the low correlation between encoding and memory measures is the result of the low value of Path B (i.e., the relation between the latent encoding and memory variables). Second, Shweder and D'Andrade assume that the high correlation between M and S is produced by the direct Path C rather than the indirect paths (AB) mediated by encoding. That is, rather than semantics determining the relations between observed behavior at the encoding phase, they assume that semantics "fill in" these relations at the later reconstruction or memory phase. Furthermore, because M and S are highly correlated but neither E and S nor E and M are, the conclusion seems to follow that Paths A and B are small. To be certain of this conclusion, however, we must be con-

² Shweder and D'Andrade seem to believe that ongoing behavior can be accurately encoded, so error in perceiving behavior is not relevant to their hypothesis.

vinced that encoding is measured as validly as the other variables.

Measurement of Behavior Encoding

The measurement of ongoing behavior usually involves a set of behavior categories, a sampling scheme for observing behavior and a coding scheme for recording the behavior that occurs (cf. Altmann, 1974; Sackett, 1978). Because the major function of behavioral observation is to determine behavior rates, the recording scheme that is usually used is an identification code with which an observer assigns a score of one to the most salient behavior categories observed in any period and a score of zero to all other categories. Frequency counts for each behavior are then obtained by summing the ones and zeroes over observations. Shweder and D'Andrade (1979) recommended the use of this coding because of its simplicity and face validity: "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).

Such simple coding schemes can be contrasted with more differentiated codes that are seldom used in observation research. The ideal procedure would involve a set of observers, each of whom observes only one behavior category for each subject of observation. Although each observer would monitor the same situation (either in person or from video recordings), a continuous record of the subject's behavior would be recorded on a scale ranging from a high to a low degree of the behavior in question (perhaps with a pen that traces the observer's judgment on a roll of steadily moving paper). This record would then be independent of the semantic biases of other observers. In addition, the continuous records could be averaged to obtain rate information or they could be correlated with the records of other categories to determine how categories covary. Although this procedure is not only ideal but perhaps utopian, there is one aspect that can easily be adapted for use in research, namely, the scaling of observations along each of the categories in the behavior inventory.

To compare scaling over all categories with identification coding, we consider a simple

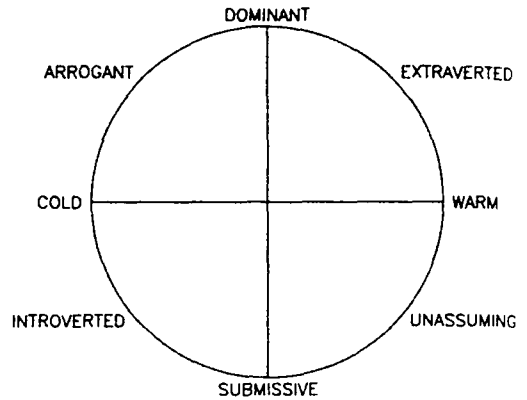


Figure 2 Wiggins's circumplex model of interpersonal trait terms.

two-dimensional category system that has been studied by several researchers (e.g., D'Andrade, 1965; Leary, 1957). According to a recent analysis by Wiggins (1979), the same dimensions (dominance-submission and warmth-coldness) form a circumplex structure underlying a wide assortment of interpersonal traits. As shown in Figure 2, eight of these traits are defined by the four poles of the trait dimensions and by the four locations intermediate to these poles. Correlations between all possible pairs of these categories are shown in Table 2. These correlations indicate that neighboring categories are highly correlated ($r = .71$). For example, although dominance and extraversion are distinguishable behaviors, they are both semantically related to the underlying dominance dimension. Thus behavior categories may have considerable semantic overlap even though they are descriptive of different ranges of behavior.

When any of the categories in the system is observed in behavior, it can be uniquely identified by using a simple recording scheme. Nevertheless, the use of identification coding obscures the implicit semantic relations between the categories. As an example, identification codings for the eight categories are shown in Table 3. Each category is observed once and the corresponding frequency count is recorded (as one for the observed behavior and zero for the rest). The correlation matrix between these observations for four categories is also shown in the table. The values are uniformly negative and would approach zero

Table 2
Correlation Coefficients Between Traits in Wiggins's Circumflex Model of Trait Terms

Trait	1	2	3	4	5	6	7	8
1. Dominant	—							
2. Arrogant	.71	—						
3. Cold	.00	.71	—					
4. Introverted	-.71	.00	.71	—				
5. Submissive	-1.00	-.71	.00	.71	—			
6. Unassuming	-.71	-1.00	-.71	.00	.71	—		
7. Warm	.00	-.71	-1.00	-.71	.00	.71	—	
8. Extraverted	.71	.00	-.71	-1.00	-.71	.00	.71	—

as the number of categories in the coding system was increased.

The corresponding codes that might be obtained if we used scaling over all categories is also shown in the table. Using a 5-point scale, we assign a value of 2 to the category most descriptive of the behavior, a value of 1 to the categories closest on the circumplex, and so on down to -2 . When these scalings are correlated between the categories, a matrix (shown in the table) that is reflective of the circumplex structure results. Comparing the semantically veridical correlations with the ones obtained from identification coding, we see that identification coding underestimates relations between semantically similar categories and overestimates relations between semantically dissimilar categories.

The bias introduced by identification coding does not disappear merely as a result of aggregation over repeated observations of the same individuals. To the extent that individuals show stability in their behavior, the bias we describe will remain. At the other extreme, if individuals exhibited no stability, the bias would disappear. The most likely empirical consequence of aggregation, however, is the cancellation of measurement error and the increased ability to observe stability in individual behavior (Epstein, 1979). Thus to the extent individuals show any stability in their behavior, aggregation over observations would be expected to increase the reliability of frequency counts and hence to maintain rather than to diminish the bias.

Block et al. (1979) noted that "the use of frequency counts of behavior is not a sufficient means of operationalizing complex psychological concepts" (p. 1062). However, they did not pursue the implications of identifica-

tion coding for determining the relations between behavior categories. In response Shweder and D'Andrade (1979) stated that the criticism

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 occurred? (p. 1079)

Our point is that identification coding does not allow the true semantic relations between behavior categories to emerge even if they are reliably exhibited in behavior. In consequence, the evidence for systematic distortion may be entirely attributable to the use of identification coding in the recording of observations and the use of scaling over all categories in subsequent memory-based ratings. Because the examples presented in Table 3 involve errorless data, they do not provide a sufficient demonstration of the distortion that may be obtained with real data. There are several ways in which to show the effect of this distortion. One would be to simulate the behavior of raters (using Monte Carlo techniques) and to observe the effect of error on the results of the two coding systems. Another would be to compare the results of the two coding systems with two different sets of raters observing the same set of actors.

Because we are more concerned with the potential distortion that is due to identification coding than we are with the issue of what behaviors are observed or what traits are inferred from observations of behavior,

Table 3
Comparison Between Two Coding Schemes for Eight Observations of the Wiggins's Circumplex

Behavior	Codings								Correlations															
	Identification				Scaling				Identification				Scaling											
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	D	E	W	S	D	E	W	S
Dominant	1	0	0	0	0	0	0	2	1	0	0	-1	-2	-1	0	1	-	-	-	-	-	.67	-.67	-.67
Extraverted	0	1	0	0	0	0	0	1	2	1	0	0	-1	-2	-1	0	-.14	-	-	-	-.14	-.67	-.67	-.67
Warm	0	0	1	0	0	0	0	0	0	1	2	1	0	-1	-2	-1	-.14	-.14	-	-	-.14	.00	.67	-
Unassuming	0	0	0	1	0	0	0	-1	0	1	0	2	1	0	-1	-2	-	-	-	-	-	-	-	-
Submissive	0	0	0	0	1	0	0	-2	-1	0	1	2	1	0	-1	0	-	-	-	-	-	-	-	-
Introverted	0	0	0	0	0	1	0	-2	-2	-1	0	1	2	1	0	-	-	-	-	-	-	-	-	
Cold	0	0	0	0	0	0	1	0	-1	-2	-1	0	1	2	1	-	-	-	-	-	-	-	-	
Arrogant	0	0	0	0	0	0	0	1	0	-1	-2	-1	0	1	2	-	-	-	-	-	-	-	-	

Note D = dominant, E = extraverted, W = warm, and S = submissive.

we presented observers with hypothetical actors whose behavior was constrained to reflect one behavior category per situation. Because of the conceptual nature of this demonstration we collected data from only four judges in each of our two conditions. To approximate the effects of aggregation, we further constrained each actor's behavior to reflect a high degree of trait stability across situations. A comparison between the two coding conditions permits us to demonstrate that identification coding produces the typical pattern of systematic distortion but that scaling over all categories retains the semantic relations between categories at both the encoding and memory phases of judgment.

Method

In this experiment, observers judged instances of behavior either by identifying the best description or by scaling the instances across all categories. Wiggins's (1979) circumplex model of interpersonal trait terms was used to define an inventory of eight behavior categories D'Andrade (1965) analyzed a similar circumplex in this first statement of the systematic distortion hypothesis. Four examples of each behavior category were selected from Wiggins's final set of adjective scales. Each set of four examples was arbitrarily assigned to one of eight hypothetical actors. For example, instances of "dominant" are *self-confident*, *domineering*, *assertive*, and *firm*. We converted each adjective into an observation by describing an instance of an actor's behavior, using the adjective. For example, in the sentence "Rick was self-confident at the meeting," *self-confident* is an observation of Rick. All observations were similarly worded in that an actor (e.g., Rick) was seen doing something (e.g., being self-confident) in a certain situation (e.g., at the meeting). The four replications of each of the eight circumplex poles resulted in 32 observations.

Procedure

The observers' task was (a) to form an impression of each of the eight actors, (b) to record the 32 observations using one of two coding schemes, and (c) to rate from memory each of the actors on each of the eight circumplex traits. Observers were told that they would be given brief descriptions of behavior of eight different people. Their task was to try to form an impression of what each person was like. Furthermore, they were asked to rate each description. In the identification condition, their rating was to be made by indicating which of the eight circumplex traits best described each behavior. In the scaling condition, they were asked to rate each description along each of the eight circumplex traits. They made the ratings using an equal-interval scale ranging from 1 (*not at all*) to 7 (*very much*).

Preliminary testing revealed that the memory task was difficult for our observers. To facilitate recall of the actors'

Table 4
Results of Experiment for Each Observer

Observer	$\bar{\alpha}$	Correlations between							
		Immediate coding and memory rating ^a	Immediate and memory matrices ^b		Immediate and semantics matrices ^b		Memory and semantics matrices ^b		
			<i>r</i>	<i>r_s</i>	<i>r</i>	<i>r_s</i>	<i>r</i>	<i>r_s</i>	
Identification condition									
A	.93	.66	.31	.26	.19	.27	.74	.79	
B	.84	.30	.43	.22	.43	.33	.73	.74	
C	.82	.65	.53	.58	.49	.49	.67	.69	
D	.79	.40	.52	.57	.49	.48	.53	.57	
<i>M</i>	.85	.50	.45	.41	.40	.39	.67	.70	
Scaling condition									
E	.95	.82	.99	.98	.73	.72	.74	.70	
F	.95	.95	.99	.96	.71	.74	.73	.73	
G	.91	.11	.92	.86	.76	.79	.62	.73	
H	.81	.36	.74	.72	.65	.74	.43	.45	
<i>M</i>	.91	.56	.91	.88	.71	.75	.63	.65	

Note Correlations between immediate, memory and semantics matrices are reported with Pearson (*r*) and Spearman (*r_s*) coefficients.

^a *N* = 64. ^b *N* = 28.

behavior, observers were first asked to learn the eight actors' names.³ When subjects could recall all the names, the experiment began. Behavior descriptions were presented on a small television monitor controlled by an Apple II computer. The computer was programmed to randomly assign each of the actors to one of the eight poles of the circumplex. It was also programmed to order the 32 behavior descriptions in a block random order. Observers responded by typing their answers directly into the computer terminal, and so no record of their judgments was available to them. Observers could take as long as needed to inspect each behavior description before indicating their response.

Immediately following the encoding phase of the experiment, observers were asked to rate each actor on each of the circumplex traits. The order for rating the actors and traits was random across subjects. For example, the actor was presented in the question, "How _____ was Rick?" in which the blank was filled with one of the eight traits. Observers could take as long as needed to respond before the next rating was requested. They again recorded their ratings by typing a number between 1 and 7 into the terminal.

Subjects

Eight graduate psychology students volunteered to participate in the study. They were randomly assigned to either of the coding conditions with four in each. They were unaware of the purpose and issues of the research at the outset.

Results

During the encoding phase of the experiment, observers were shown each actor four

times. This procedure allowed us to estimate the reliabilities of the respective coding schemes. To do this, we calculated coefficient alpha for each behavior category for each observer.⁴ Because these coefficients were uniformly high for all eight categories, we report each observer's average α for the eight categories. As shown in Table 4, the reliabilities were equally high for both schemes. Thus any differences between these codings are unlikely to be attributable to the consistency of responses with each coding instrument. Because these reliabilities were high, we summed each observer's judgments over the four replications to obtain a single Target \times Behavior (8 \times 8) encoding matrix.

To determine how well observers' memory-based impressions matched their immediate encodings, we calculated the correlation between their encoding scores and their mem-

³ The names we selected for the eight targets were Don, Doug, Fred, Jack, Jerry, Rick, Roy, and Walter.

⁴ We found α by treating the four blocks of observations as analogous to four items, and the eight actors as analogous to individuals responding to these items. That is, α is an estimate of the generalizability of between-actor differences based on the variation between actors within blocks and the variation within actors across blocks.

ory-based ratings. These correlations indicate that observers forgot some of their immediate impressions between the two parts of the experiment (Table 4). Indeed, one observer's (G's) correlation, $r(62) = .11$, failed to reach significance. Such forgetting is necessary for systematic distortion to appear. Thus our experimental analog of behavioral observation satisfies the requirements for the typical demonstration of systematic distortion.

Examination of the (Pearson and Spearman) correlations between the semantic structure matrix (i.e., the theoretical circumplex pattern in Table 2) and the matrices of Pearson correlations between behavior categories for both phases of the experiment enables us to determine the role of semantics in judgment. As shown in Table 4, the relations among semantics, encoding, and memory depended on the coding scheme. The identification condition replicated the typical systematic distortion pattern. The semantic and memory matrices were more highly correlated (average $r = .67$) than the semantic and encoding matrices (average $r = .40$). Thus semantics were more closely related to memory ratings than to encodings, and the latter were only moderately related to memory ratings (average $r = .45$). Spearman coefficients yielded the same results.

In the scaling condition, however, we found that the semantic matrices were no more highly related to the memory matrices (average $r = .63$) than to the encoding matrices (average $r = .71$). Indeed, the relations between semantics and encoding were slightly larger than between semantics and memory. Also, we saw that the relations between behavior categories at the encoding and memory phases were quite similar (average $r = .91$), so that little changed in the relations between categories in the two phases of the experiment. It appears therefore that the typical systematic distortion pattern does not obtain when scaling is used to record observations.

The most interesting comparison, however, was the one revealing the striking difference between the identification and scaling conditions for the relations between immediate and semantic matrices. In the identification condition, the immediate-semantic correlations ranged from .19 to .49, whereas in the scaling condition they ranged from .65 to .76.

This suggests that semantics are highly related to immediate encodings of behavior, but only when the appropriate measurement is taken.

In Table 5 we illustrate some of the differences between the results of the two coding systems (behavior correlations for observers A and E)⁵ and we illustrate how the patterns predicted earlier (Table 3) were obtained. In the identification condition (Observer A), Pearson correlations between behavior categories in the immediate encodings tended to be negative and near zero. The average absolute correlation was only .17. The average absolute correlation for A's memory-based ratings was much larger (.63), which could suggest that systematic distortion has occurred. That this was a consequence of the coding scheme rather than biases in recall, however, is seen in the results of the scaling condition (Observer E). E's average absolute correlations between behavior categories were large and similar at both encoding (.63) and memory (.59) stages. Furthermore, the individual correlations were more evenly balanced between positive and negative scores. There is thus evidence for systematic distortion only when the observer's ratings are limited by the use of identification coding. When ratings are used in both the encoding and memory phases of the experiment, the typical systematic distortion pattern disappears.

Discussion

The purpose of this research was to demonstrate the importance of coding schemes for recording the frequency of observed behavior. We argued that frequency counts of simple behavior identifications introduce a method bias that attenuates the validity of behavior observations. As a result, the absolute magnitudes of correlations between behavior categories are attenuated.

Memory-based ratings of behavior frequencies, however, are typically scaled across all behavior categories. These ratings tend to correlate with observed frequencies, but they also tend to correlate more strongly between behavior categories than frequencies obtained from simple identification coding. These phe-

⁵ Data for the remaining observers were similar and so were not included in the table.

Table 5
Intercorrelations Between Behavior Categories for Two Observers

Behavior	Observer A (Identification)								Observer E (Scaling)							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Immediate codings																
1. Dominant	—								—							
2. Arrogant	-.18	—							.86	—						
3. Cold	.10	-.15	—						.10	.46	—					
4. Introverted	-.18	-.15	-.15	—					-.96	-.76	.17	—				
5. Submissive	-.18	-.15	-.15	-.15	—				-.96	-.73	.19	.99	—			
6. Unassuming	-.18	-.15	-.15	-.15	-.15	—			-.93	-.69	.26	.98	.99	—		
7. Warm	-.23	-.18	-.18	-.18	-.18	-.18	—		.07	-.30	-.98	-.34	-.36	-.44	—	
8. Extraverted	-.24	.18	-.19	-.19	-.19	-.19	.04	—	.91	.70	-.26	-.99	-.98	-.97	.40	—
Memory Ratings																
1. Dominant	—								—							
2. Arrogant	.86	—							.90	—						
3. Cold	.63	.66	—						.32	.51	—					
4. Introverted	-.89	-.70	-.40	—					-.88	-.78	.04	—				
5. Submissive	-.91	-.79	-.55	.79	—				-.88	-.78	.04	.99	—			
6. Unassuming	-.91	-.91	-.53	.81	.71	—			-.86	-.74	.11	.96	.96	—		
7. Warm	-.54	-.71	-.80	.19	.41	.53	—		-.01	-.31	-.82	-.34	-.34	-.34	—	
8. Extraverted	.65	.54	.03	-.87	-.65	-.65	.13	—	.91	.70	-.05	-.88	-.88	-.93	.26	—

nomena were replicated in the present experiment. Ever since Newcomb (1931) first noted these facts, researchers have questioned the validity of behavior ratings. Shweder and D'Andrade (1979) formulated this skepticism in terms of a systematic distortion that is introduced by observers' reliance on semantic relations between behavior categories rather than their actual co-occurrence. Our point is that it is the validity of behavior coding that is in question and that the true interbehavior relations are more validly measured with ratings.

Thus in the present experiment, relations between immediate ratings of behavior correlated highly with relations between memory-based ratings ($r = .91$). When identification was used to measure encoding, however, the relations between behavior categories did not correlate as highly with relations between memory-based ratings ($r = .45$). These results indicate that it may not be anything peculiar to memory-based ratings that introduces stronger correlations between behavior categories; rather, it is the use of ratings per se that is responsible, whether they are conducted immediately or from memory. Thus Shweder and D'Andrade's (1979) point that semantic relations predict memory relations better than immediate encoding relations may also be a function of coding method. We replicated the systematic distortion in memory phenomenon with identification coding; however, the phenomenon disappeared when observations were scaled across categories at both encoding and memory phases. Indeed, semantic relations predicted immediate encodings slightly better than they predicted memory relations.

Whereas Shweder and D'Andrade (1979) interpreted the correlation between semantics and memory as causal (Path C in Figure 1), the present results with ratings suggest an alternate hypothesis. According to our model, the relation between semantics and memory can be completely predicted by the effect that semantics have on encoding (Path A) and the effect of encoding on memory (Path B). We would expect the correlation between semantics and memory to be equal to the product of our estimates of Paths A and B (.71 and .91). From our results, this product (.65) agrees well with the obtained correlation

Table 6

Intercorrelation Matrices Derived From Immediate Encodings of Behavior (Upper Triangle) and Subsequent Ratings (Lower Triangle) in Newcomb's (1931) Study

Behavior	1	2	3	4	5
1.	—	.52	.05	.29	.20
2.	.67	—	.03	-.14	.08
3.	.61	.68	—	-.11	.48
4.	.97	.88	.66	—	.16
5.	.66	.92	.77	.75	—

Note. Behavior categories are as follows: 1. Tells of his own past, and of the exploits he has accomplished. 2. Gives loud and spontaneous expressions of delight or disapproval. 3. Goes beyond only asking and answering necessary questions in conversations with counselors. 4. How is the quiet hour spent? 5. Spends a lot of time talking at the table. Data are based on observations of 30 boys at a summer camp.

(.63). Thus the present results are consistent with the hypothesis that the correlation between semantics and memory is entirely spurious (i.e., $C = 0$) in that the correlation is the result of semantics at the encoding phase, an effect that is simply registered again in the memory phase. When identification coding is used, however, these conclusions do not follow. But the higher correlation between semantics and memory is shown to be the result of the fact that both of these processes are measured with ratings. Identification codings do not correlate as highly with either of these ratings as the ratings correlate with each other.

A subset of Newcomb's (1931) data, taken from Shweder (1979), illustrates the problems produced by identification coding. Newcomb's observers recorded various instances of behavior reflecting extraversion-introversion among their summer-camp subjects. Only one behavior category was recorded at a time. In Table 6 we show the correlations between five of these categories as obtained from observations and subsequent ratings (on a 5-point scale for each category). As is evident, many of the observed behavior correlations are near zero or negative. The ratings correlations are much higher. Although only one of the categories might describe a given observation, they are all conceptually related. These relations cannot appear as strongly when identification coding is used as when

Table 7

Tau Coefficients Between Behavior Categories Derived From Immediate (Upper Triangle) and Memory-Based Ratings (Lower Triangle) in Shweder and D'Andrade (1980)

Category	1	2	3	4	5	6	7	8	9	10	11
1. Agree	—	<i>-.67</i>	<i>.00</i>	<i>.33</i>	<i>.00</i>	<i>-.67</i>	<i>-.67</i>	<i>.00</i>	<i>1.00</i>	<i>.00</i>	<i>.33</i>
2. Comply	<i>.28</i>	—	<i>.33</i>	<i>-.67</i>	<i>-.33</i>	<i>.33</i>	<i>-.33</i>	<i>-.33</i>	<i>-.67</i>	<i>.33</i>	<i>.00</i>
3. Praise	<i>.29</i>	<i>.39</i>	—	<i>-.67</i>	<i>.33</i>	<i>-.33</i>	<i>.33</i>	<i>-1.00</i>	<i>.00</i>	<i>.33</i>	<i>-.67</i>
4. Advise	<i>.07</i>	<i>.07</i>	<i>.08</i>	—	<i>.00</i>	<i>.00</i>	<i>.00</i>	<i>.67</i>	<i>.33</i>	<i>-.67</i>	<i>.33</i>
5. Inform	<i>.11</i>	<i>.11</i>	<i>-.05</i>	<i>.42</i>	—	<i>-.33</i>	<i>.33</i>	<i>-.33</i>	<i>.00</i>	<i>-.33</i>	<i>-.67</i>
6. Suggest	<i>.06</i>	<i>.12</i>	<i>.03</i>	<i>.51</i>	<i>.37</i>	—	<i>.33</i>	<i>.33</i>	<i>-.67</i>	<i>-.33</i>	<i>.00</i>
7. Question	<i>-.01</i>	<i>-.11</i>	<i>-.06</i>	<i>.24</i>	<i>.14</i>	<i>.11</i>	—	<i>-.33</i>	<i>-.67</i>	<i>-.33</i>	<i>-.67</i>
8. Criticize	<i>-.31</i>	<i>-.31</i>	<i>-.46</i>	<i>.10</i>	<i>.00</i>	<i>.13</i>	<i>.12</i>	—	<i>.00</i>	<i>-.33</i>	<i>.67</i>
9. Disagree	<i>-.42</i>	<i>-.18</i>	<i>-.31</i>	<i>.05</i>	<i>-.10</i>	<i>-.02</i>	<i>-.01</i>	<i>.59</i>	—	<i>.00</i>	<i>.33</i>
10. Joke	<i>.16</i>	<i>-.25</i>	<i>.21</i>	<i>-.33</i>	<i>-.21</i>	<i>-.49</i>	<i>-.19</i>	<i>-.30</i>	<i>-.11</i>	—	<i>.00</i>
11. Ridicule	<i>-.24</i>	<i>-.50</i>	<i>-.27</i>	<i>-.03</i>	<i>-.24</i>	<i>-.29</i>	<i>-.13</i>	<i>.39</i>	<i>.45</i>	<i>.17</i>	—

Note: Numbers in italics represent correlations between categories that are semantically similar

ratings are used. Thus Newcomb's conclusion that the ratings contain semantic biases may not be an appropriate interpretation.

In a more recent study by Shweder and D'Andrade (1980), three observers identified instances of interpersonal behavior among four videotaped family members. The categories were examples of normal kinds of verbal behavior (e.g., agree, comply, praise). The resulting matrices of τ coefficients for encoding and memory are shown in Table 7. Many of the 55 τ coefficients (23) they reported for encoding relations among 11 of these categories were less than $-.30$. The corresponding coefficients for memory-based ratings (on a 1-7 scale) contained only 9 categories that were less than or equal to $-.30$. The difference between matrices in the occurrence of extreme negative values is significant ($z = 3.84, p < .05$).

One interesting aspect of this study is that Shweder and D'Andrade (1980) did not restrict their observers' encoding responses to only one category in the behavior inventory. Thus the bias produced by uniquely identifying behavior by a single category may have been attenuated. Nevertheless, we do not know how carefully observers were instructed to use all categories that are semantically related to an actor's behavior. It is clearly not necessary to use all relevant categories to accurately describe behavior. However, if observers actually used all the categories that

were semantically related to an observation, we would expect semantically similar categories to be positively correlated. That this was not the case can be seen by inspecting the correlations between categories that are semantically similar. These correlations are arrayed in italics along the diagonal of the matrices. The average τ correlation between these categories in the encoding matrix is $-.17$, whereas the corresponding value in the memory matrix is $.31$. These averages are significantly different, $t(9) = 24.0, p < .05$. For example, the correlation between encodings of *advise* and *suggest* is $.00$, whereas the correlation based on memory ratings is $.51$. If observers used only one of these semantically related categories at a time to describe either form of behavior, the low correlation might obtain. Ratings, however, might reflect the high degree of semantic overlap in these categories. It appears, therefore, that the bias produced by identification coding may have contributed to the poor match they obtained between encoding and memory matrices ($r = .22$). Thus their interpretation that ratings contained semantic biases may also not be appropriate.

Researchers (e.g., D'Andrade, 1965; Shweder, 1975) have found that semantic relations between items in personality questionnaires also are predictive of the obtained correlations between the items when they are used in personality assessment. There are at least two

ways to interpret this finding. One advanced by Shweder and D'Andrade (1979, 1980) is that these correlations are illusory in that the actual behavior of the measured individuals does not contain the obtained structure. Instead, the ratings are said to be biased by inferred semantic relations between items or by what some writers have termed an implicit personality theory (Bruner & Tagiuri, 1954; Schneider, 1973).

Another interpretation is that this relation demonstrates the validity of the personality test. Because raters are using the items in a semantically consistent manner, their scores display internal consistency. Physical scientists who measured the two sides of various rectangles would expect these ratings to be orthogonal if sufficient variation in the variables were observed. This structure would be implied by the semantic independence of these concepts. If such independence were not obtained, something would be considered amiss in the scientists' instruments. If the scientists also found this pattern in their memory for their observations, they would be encouraged that their memories displayed consistency. Whether their memories of specific lengths and widths were accurate would, of course, be another question.

We think that the present results support this interpretation. According to our model of behavior encoding, semantics define the relations between events (e.g., behavior) and behavior categories. Because any behavior has a complete set of semantic relations with every category, semantics partly define the structure of items in personality tests. The present results indicate that semantics can correlate as highly with immediate observations as with memory-based impressions and that the ability to recover these relations at the memory phase is entirely explained by the role that semantics play at the encoding stage. Furthermore, these findings appear to be independent of the accuracy of the observer's memory. Therefore, to the extent that scales are used correctly, one should expect semantics to partly recover the empirical relations between scales.

The latter conclusion rests, however, on the assumption that individuals differ in stable ways on some observable dimensions. The existence of individual differences has not

been questioned by Shweder and D'Andrade (1979, 1980). Furthermore, skepticism about this issue seems to be answered in part by Epstein's (1979) findings that behavioral stability is available for observation if enough samples of behavior are measured. Thus the assumption that individual differences exist (e.g., some people are more friendly than others) seems to imply that trait relations will follow semantics (friendly individuals will tend to be less hostile).

This conclusion is relevant to one of the earliest reports of evidence for systematic distortion (Newcomb, 1931). In that study, Newcomb found that behavior ratings correlated significantly (average $r = .41$) with observed behavior frequencies (that were identification coded). What puzzled Newcomb was that the interobserver reliabilities of the ratings were as high for behavior categories (.64) that were observed by only one or none of the raters (e.g., making one's bed) as for categories (.71) that were observed by many (e.g., lying around):

Behaviors never recorded, never seen, and those felt to be highly uncertain were as uniformly rated as those at the opposite extremes. Since the guessed ratings are of highly questionable validity, must not the others, which are no more uniform, be almost equally invalid? (p. 289)

An alternative explanation rejected by Newcomb (1931) is that observers used instances of observed behavior to infer the behavior of subjects in unobserved contexts. If people do exhibit some stability in their behavior and they differ in personality, then predictions to unseen but conceptually related behavior would seem to be a rational procedure (cf. Ajzen, 1977). In the absence of data (which Newcomb did not provide), we cannot say whether such predictions are valid. However, the present analysis and results suggest that the presence of prediction to unobserved behavior does not imply that the ratings of actually observed behavior are invalid.

It may seem to some readers that the semantic interpretation of behavior ratings we present here logically implies the existence of traits insofar as individuals differ in some personality dimension. However, most reported observations of behavior have been conducted over only a limited set of situations. As Shweder and D'Andrade (1979) note, the

trait concept also implies cross-situational consistency. Mischel (1973) makes this criterion a major focus of his critique. His point is that such consistency is limited. Thus even if traits seem to be an inevitable outcome of individual differences, the possibility remains that such consistencies are eliminated when individuals are observed across different situations.

If this is indeed the case, then we would need models of personality functioning that could predict why people behave differently across situations. Such models would presumably specify how individual characteristics (traits?) interact with situations. Note, however, that this type of theorizing still requires the postulation of stable individual characteristics that can function as terms in the person-situation interaction. Examples of this type of theorizing are evident in the work of Atkinson (1957, 1964, 1978) in the domain of achievement motivation and of Eysenck (1967, 1976, 1981) in the domain of introversion-extraversion. A further example may be found in the studies by Revelle and his associates (Humphreys, Revelle, Simon, & Gilliland, 1980; Revelle, Amaral, & Turriff, 1976; Revelle, Humphreys, Simon, & Gilliland, 1980), who have shown that the personality trait of impulsivity has systematic, although complex relationships to cognitive performance. These models attempt to explain behavior as a joint function of both stable individual characteristics and situational variables that are defined in a theoretical model of sufficient complexity to make predictions across situations and individuals. This interaction approach would seem to be profitable for understanding personality.

Although the present experiment only approximates what we regard as an ideal observation procedure, it does show that use of ratings at the encoding phase is a critical factor in estimating the true relations between immediately observed behavior categories. Our use of behavior descriptions rather than actual physical events was a helpful shortcut for demonstrating this point. If we had presented the actual behavior of our actors, the encodings our observers would draw might have lower reliability, but this would be little reason to expect this factor to differentially affect the reliability of the two coding schemes.

Indeed, the reliabilities we obtained for these schemes were equally high, a finding that supports our contention that it is not reliability of these codes but rather their validity that does or does not produce support for systematic distortion.

Our purpose in presenting this research is not to argue that distortion never occurs. There are several findings that show some evidence of bias in observer reports (e.g., Berman & Kenny, 1977; Hamilton & Rose, 1980). Although these findings may lack external validity (Block, 1977), our point is that such distortions are not necessarily so large that valid assessments cannot be obtained. Furthermore, we feel that distortion is possible at any stage in the observation and recall of behavior and is not necessarily confined to memory-based ratings. Shweder and D'Andrade (1979, 1980) argued that immediate encodings have validity that no other methods possess. Distortion at the encoding stage has been observed, however (cf. Pettigrew, 1979), and there are often theoretical reasons for expecting such distortion (e.g., in intergroup conflict and race relations). Because behavior takes its meaning only by virtue of its encoding, disagreement about its meaning would seem inevitable. Nevertheless, methods are available for reducing such errors with such simple devices as averaging over observers (Kenny & Berman, 1980).

What we feel is unjustified, however, is the conclusion that observers cannot draw correlational inferences from their experience. Neither the present research nor any we are aware of justifies this conclusion. Future researchers of encoding and memory of behavior should perhaps focus on how both objective characteristics of behavior and the cognitive-motivational processes of observers combine to yield the encodings we commonly produce. Our analysis and results suggest that such models should emphasize semantics as a critical mediator in this process.

References

- Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, 35, 303-314.
- Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behavior*, 69, 227-267.
- Atkinson, J. W. (1957). Motivational determinants of

- risk-taking behavior. *Psychological Review*, 64, 359-372.
- Atkinson, J. W. (1964). *An introduction to motivation*. New York: Van Nostrand.
- Atkinson, J. W. (1978). Strength of motivation and efficiency of performance. In J. W. Atkinson & J. Raynor (Eds.), *Personality, motivation and achievement* (pp. 117-142). Washington, DC: Hemisphere.
- Berman, J. S., & Kenny, D. A. (1977). Correlational bias: Not gone and not to be forgotten. *Journal of Personality and Social Psychology*, 35, 882-887.
- Block, J. (1977). Correlational bias in observer ratings: Another perspective on the Berman and Kenny study. *Journal of Personality and Social Psychology*, 35, 873-880.
- Block, J., Weiss, D. S., & Thorne, A. (1979). How relevant is a semantic similarity interpretation of personality ratings? *Journal of Personality and Social Psychology*, 37, 1055-1074.
- Bowers, K. S. (1973). Situationism in psychology: An analysis and critique. *Psychological Review*, 80, 307-336.
- Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634-654). Cambridge, MA: Addison-Wesley.
- Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, 90, 272-292.
- D'Andrade, R. G. (1965). Trait psychology and componential analysis. *American Anthropologist*, 67, 215-228.
- D'Andrade, R. G. (1974). Memory and the assessment of behavior. In T. Bialock (Ed.), *Measurement in the social sciences* (pp. 159-186). Chicago: Aldine-Atherton.
- Endler, N. S., & Magnusson, D. (1976). *Interactional psychology and personality*. Washington, DC: Hemisphere.
- Epstein, S. (1979). The stability of behavior. I On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37, 1097-1126.
- Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Charles C Thomas.
- Eysenck, H. J. (1976). *The measurement of personality*. Baltimore, MD: University Park Press.
- Eysenck, H. J. (Ed.) (1981). *A model for personality*. New York: Springer-Verlag.
- Fiske, D. (1978). *Strategies for personality research: The observation versus interpretation of behavior*. San Francisco: Jossey-Bass.
- Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, 39, 832-845.
- Humphreys, M. S., Revelle, R., Simon, L., & Gilliland, K. (1980). Individual differences in diurnal rhythms and multiple activation states: A reply to M. W. Eysenck and Folkard. *Journal of Experimental Psychology: General*, 109, 42-48.
- Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, 34, 107-117.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kenny, D. A., & Berman, J. S. (1980). Statistical approaches to correction for correlational bias. *Psychological Bulletin*, 88, 288-295.
- Leary, T. (1957). *Interpersonal diagnosis of personality—A functional theory and methodology for personality evaluation*. New York: Ronald Press.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 80, 252-283.
- Newcomb, T. (1931). An experiment designed to test the validity of a rating technique. *Journal of Educational Psychology*, 22, 279-289.
- Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, 5, 461-476.
- Revelle, W., Amaral, P., & Turriff, S. (1976). Introversion/extraversion, time stress, and caffeine: The effect on verbal performance. *Science*, 192, 149-150.
- Revelle, W., Humphreys, M. S., Simon, L., & Gilliland, K. (1980). The interactive effect of personality, time of day, and caffeine: A test of the arousal model. *Journal of Experimental Psychology: General*, 109, 1-31.
- Sackett, G. P. (1978). *Observing behavior* (Vol. 2). Baltimore, MD: University Park Press.
- Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, 79, 294-309.
- Shweder, R. A. (1975). How relevant is an individual difference theory of personality? *Journal of Personality*, 43, 455-484.
- Shweder, R. A. (1977). Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. *Current Anthropology*, 18, 637-658.
- Shweder, R. A. (1979). Rethinking culture and personality theory, Part I: A critical examination of two classical postulates. *Ethos*, 7, 255-278.
- Shweder, R. A., & D'Andrade, R. G. (1979). Accurate reflection or systematic distortion? A reply to Block, Weiss, and Thorne. *Journal of Personality and Social Psychology*, 37, 1075-1084.
- Shweder, R. A., & D'Andrade, R. G. (1980). The systematic distortion hypothesis. In R. Shweder (Ed.), *New directions for methodology of social and behavior science* (Vol. 4, pp. 37-58). San Francisco: Jossey-Bass.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29.
- Wiggins, J. S. (1979). A psychological taxonomy of trait terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37, 395-412.

Received February 28, 1983
Revision received June 27, 1983 ■