

Some Determinants of Factor Structures From Personality-Trait Descriptors

Dean Peabody
Swarthmore College

Lewis R. Goldberg
University of Oregon and Oregon Research Institute

Three determinants of the factor structures of personality traits are investigated. The 1st, selection of variables, was controlled by using 57 bipolar scales, selected to be representative of common trait terms. In analyses of 7 data sets, variants of the "Big Five" factors were always found. Factor similarities were very strong for the 3 largest factors, which were transformed into general evaluation and 2 descriptive dimensions. As a 2nd determinant, judgments about real people were compared with judgments about the conceptual relations among traits. Factor structures based on the 2 types of judgments are similar, but those based on conceptual judgments tend to be simpler. The 3rd determinant involved the degree of restriction of the sample to evaluatively homogeneous targets. Restriction of range reduced the size of all factors, especially Factor II. Findings from previous studies are integrated within this framework.

The variety of personality characteristics is enormous. At the same time, there are many relations among these traits. The first fact has encouraged the major activity of personality psychologists: the conception and measurement of particular personality characteristics. The second fact has encouraged some psychologists to try to summarize the interrelations and to look for a structural representation of all personality traits. A strong case can be made that any particular personality characteristic can best be interpreted in the context of a comprehensive structure.

In the quest for such a personality-trait structure, one must decide at least implicitly on the domain of characteristics to study and on the method to be used to summarize the relations among these characteristics. Of the myriad ways that humans differ from one another, we assume that the most important have eventually become encoded in our language as trait-descriptive terms (e.g., *bold, warm, thorough, relaxed, intelligent*). By far the most popular methodology for summarizing the relations among personality characteristics has been factor analysis, here used in the broadest sense of that technique. In this article, we consider both the similarities and the differences in the personality factors so derived. In a form of "experimental

psychometrics," we apply the experimental method to predict as dependent variables both the size and the content of the factors. We have studied as independent variables three influences on these factors.

Specifically, we assume that by far the most important influence on any factor structure is the selection of variables under study. In the first section of this article, we provide a rationale for the selection of a particular set of personality variables, a set that is designed to be representative of the domain of English trait adjectives. In the second section, we describe and illustrate two other potential influences on structural representations. The first of these is the nature of the data used in the analyses: (a) internal judgments of the conceptual relations among the traits or (b) external judgments of the extent to which the traits are descriptive of actual people. In analyses of external judgments, another influence on factor structures is the degree of "restriction of range" of the raters' evaluations of the targets, specifically whether the target sample includes both people who are liked and people who are not.

In the third section of this article, we present findings from seven data sets that were specifically designed to estimate the relative effects of these two influences on the resulting factor structures. In each case, the results are based on the same selection of variables and the same factor analytic procedures. Our findings demonstrate that five similar factors emerge within each of the data sets, although their sizes and locations are affected systematically by both of the influences under study.

In the fourth section, we review those previous studies that have included a relatively comprehensive coverage of personality-trait adjectives, and we examine the congruence between their findings and our own. In so doing, we confirm the effects of the two influences we have identified. In addition, we show that some of the major differences can be traced to the particular sets of variables selected in each study. In the final section, we discuss some of the implications of our analyses, including the advantages of a standard structural representation for personality traits.

This project was supported by National Institute of Mental Health Grant MH-39077 and a Faculty Research Grant from Swarthmore College.

We are enormously indebted to those colleagues who helped us improve the clarity of this article, including David M. Buss, Robyn M. Dawes, John M. Digman, Hans J. Eysenck, Donald T. Fiske, Sarah E. Hampson, Robert Hogan, Oliver P. John, Henry F. Kaiser, Kevin Lanning, Robert R. McCrae, Warren T. Norman, Leonard G. Rorer, James A. Russell, Lawrence J. Stricker, Auke Tellegen, and Jerry S. Wiggins. We are grateful to Robert R. McCrae for providing us with unpublished analyses of his data and to Loren Miller for his assistance with the data collection at Swarthmore College.

Correspondence may be addressed to Dean Peabody, Department of Psychology, Swarthmore College, Swarthmore, Pennsylvania 19081 or to Lewis R. Goldberg, Oregon Research Institute, 1899 Willamette Street, Eugene, Oregon 97401.

Rationale for Variable Selection

The selection of variables is clearly of critical importance, serving to determine the very presence of a factor, as well as its size and content. Indeed, the selection of variables is *the* method used to establish particular personality characteristics. However, is there any rationale for selecting variables to represent all personality characteristics? One such rationale derives from the assumption that the important personality characteristics will eventually become encoded in our language (see Klages, 1929/1932, and Allport, 1937). This same rationale motivated Cattell (1943) in his pioneering efforts to discover the basic factors of personality structure:

The position we shall adopt is a very direct one . . . making only the one assumption that all aspects of human personality which are or have been of importance, interest, or utility have already become recorded in the substance of language. For, throughout history, the most fascinating subject of general discourse, and also that in which it has been most vitally necessary to have adequate, representative symbols, has been human behavior. Necessity could not possibly be barren where so little apparatus is required to permit the birth of invention. (p. 483)

This rationale is not limited to psychologists or to personality-descriptive terms. For example, the philosopher Austin (1957) has noted:

our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our arm-chairs of an afternoon—the most favored alternative method. (p. 8)

For recent discussions of this rationale in the context of personality traits, see Goldberg (1981, 1982); John, Goldberg, and Angleitner (1984); and especially John, Angleitner, and Ostendorf (1988).

Historically, Allport and Odbert (1936) provided the source for a comprehensive selection of personality-related terms by listing those found in an unabridged English dictionary. By the early 1940s, Cattell had embarked on a major effort to select personality variables, with representativeness from the Allport and Odbert trait list as one of his objectives. Through a series of stages, this effort led to the construction of 35 bipolar variables (Cattell, 1947). In Cattell's own interpretation, analyses of these variables suggested at least 12 factors. However, Fiske (1949) and Tupes and Christal (1961) showed that more common methods of analysis suggested only 5 replicable factors, which have come to be referred to as the "Big Five." Further support for a 5-factor structure of the Cattell variables was provided by Norman (1963) and by Digman and Takemoto-Chock (1981). Similar 5-factor structures based on other sets of variables have been reported by Digman and Inouye (1986) and by McCrae and Costa (1985c, 1987).

Verbal labels for broad factors are hard to find, but in a general way these five factors are related to Power, Love, Work, Affect, and Intellect, respectively. For each factor, we list its roman numeral, its traditional label (in parentheses), and a "pro-

totypical" scale—the scale most highly related to that factor in our analyses:

- I (Surgency): Bold–Timid
- II (Agreeableness): Warm–Cold
- III (Conscientiousness): Thorough–Careless
- IV (Emotional Stability): Relaxed–Tense
- V (Culture): Intelligent–Unintelligent

The goal of Cattell's (1946) selection of personality traits was not solely that of representativeness. Cattell also added "all terms that have been used in technical psychological description" (1946, p. 217), many of which concerned psychopathology and so led to an overrepresentation of variables related to the Emotional Stability factor. Moreover, Cattell replaced all variables for rating aspects of intelligence with an actual intelligence test. This test was omitted from later studies, leaving no direct representation of intelligence. In its absence, a fifth factor was called *Culture*. In this article, which reports findings from studies in which ratings of aspects of intelligence have been reintroduced, we show that a more appropriate name for Factor V is *Intellect*.

From the 35 variables used by Cattell (1947), Norman (1963) selected 20, 4 marking each of the five factors identified by Tupes and Christal (1961); these 20 variables have frequently been used in later studies. However, that set of variables was not intended as a representative sampling of the trait domain. Rather, Norman selected the 4 most univocal markers of each of the five factors, thus treating the factors as five separate clusters of variables. Because variables located between the clusters were not selected, the factors tend to remain fixed in their initial positions in all later analyses of these 20 variables.

In contrast, the sole objective of our strategy for selecting variables was to achieve an adequate representation of common trait adjectives. The details of this selection are presented in Peabody (1987). A pool of 571 terms was constructed from the conjunction of several existing trait lists, and Peabody then categorized these terms according to similarity of meaning. Each resulting category could generally be paired with a contrasting category, one that included terms that were opposites both descriptively and evaluatively. To represent these paired categories, 57 antonym pairs were selected. The resulting 57 bipolar scales were then used to generate seven different data sets, described later in this article. Thus, we held constant the selection of variables, so as to be able to demonstrate the effects of other influences on factor representations.

Two Influences on Factor Size and Location

External Versus Internal Judgments

The relations among personality-trait descriptors can be obtained by either of two conceptually distinct procedures. The most direct is to try to elicit the common versus the distinct features of meaning between any pair of trait terms from the judgments of native speakers. Alternatively, these relations can be obtained indirectly from a sample of individuals' judgments about themselves or others. Wiggins (1973) used the term *external structure* to refer to a personality representation based on judgments about the characteristics of actual people, as compared with *internal structures*, which are based on judgments

of the interrelations among the characteristics themselves. Studies of personality-trait structure have traditionally relied on external judgments. In contrast, internal judgments have been used primarily in studies of person perception, where they are viewed as reflecting the perceiver's "implicit personality theory." How should the structures derived from these two types of judgments correspond?

If either type of judgment were derived from the other (or both from something else), then the two types of structural representations should be quite similar. For example, in the systematic distortion hypothesis proposed by Shweder and D'Andrade (1980), external judgments are viewed as being derived from internal judgments. Shweder and D'Andrade argued that structures based on the two types of judgments are the same. As an alternative hypothesis, we would argue from general principles of cognitive economy that internal representations should be similar to, but generally simpler and more schematic than, the environmental information that is being represented.

These tendencies—for similarity and simplification—are supported by the findings from the few previous comparisons between internal and external judgments. For example, in Peabody (1987) four judges rated the similarity of meaning on the 57 representative scales for each of the 114 separate trait terms that had been used to define the two poles of these scales. Analyses of the mean ratings of the four judges revealed factors corresponding to the Big Five as derived from external judgments, plus an additional Values factor (Honest–Dishonest, Moral–Immoral, Fair–Unfair). In size, however, three factors—II (Agreeableness), III (Conscientiousness), and I (Surgency)—were much larger than the others. Moreover, there was a tendency for each of the smaller factors (Intellect, Emotional Stability, and Values) to be only partly differentiated from the three large factors; that is, the scales most highly related to each of the smaller factors tended to have equally large loadings on one of the large factors as well. Thus, these results show some tendency toward a cognitive simplification into three factors.

These results can be interpreted as a combination of the tendencies for similarity between the two types of representations and for simplification within the internal judgments. Moreover, as it turned out each of these two tendencies was clearly illustrated in an analysis of individual judges. The four judges differed substantially in their use of the middle (neutral) category on the nine-step rating scales; two of the judges used this category for a majority of their responses, whereas the other two used it for a minority. That is, the former pair of judges recognized no degree of relatedness among a majority of traits; these ("differentiated") judges recognized relations only within limited subsets of scales, but not between the subsets. On the other hand, the latter pair of ("broad") judges reported some degree of relatedness among much larger sets of traits.

When the average ratings from the first and second pairs of judges were factor analyzed separately, the differences in the 2 factor structures were remarkably clear. For the differentiated judges, each of the 6 factors was clearly differentiated from the others; the scales most highly related to the factors of Intellect, Emotional Stability, and Values generally did not have high loadings on the first 3 factors as well. Thus, these judges provided factors that were quite similar to those found with external judgments. In contrast, analyses of the other pair of judges

revealed essentially only 3 broad factors. These broad judges show the tendency for structures derived from internal judgments to be simplified. Therefore, each of these two types of judges illustrates one of the two tendencies in internal judgments, similarity and simplification.¹

Target Likability and Restriction of Range

In general, if there is some restriction in the variation of any characteristic, then its relations with other variables will be reduced. For example, if aptitude test scores are obtained in a sample selected for high ability, the sizes of the intercorrelations among the scores (and of the factors summarizing them) are attenuated. The same principle applies to personality measures. In the studies by Cattell (1947), Tupes and Christal (1961), and Norman (1963), the members of a group described each of the others, including those whom they liked and those whom they did not like. In contrast, in many recent studies the targets have included only themselves or a close friend. With such targets, ratings tend to be restricted to the desirable portion of most scales. Consequently, in samples of liked targets one should expect smaller correlations and factors of smaller size. Moreover, such effects should vary systematically across the variables as a function of the extent to which they elicit differences between liked and disliked targets. On the basis of previous analyses of the role of evaluation on each of the Big Five factors, scales measuring aspects of Factor II (Agreeableness), such as Warm–Cold, should elicit the largest effects, and scales measuring aspects of Factor I (Surgency) should elicit the smallest effects.

Analyses of Seven Data Sets

In this section, we examine the factors derived from each of seven data sets. The selection of variables was held constant by using the same 57 scales, a selection designed to be representative of common trait adjectives. The data sets were designed to illustrate the effects of external versus internal judgments and restricted versus unrestricted range. There are five external data sets, two from self-reports and one each from liked peers, not-liked peers, and both types of peers combined. In addition, there are two internal data sets, one each from the broad and the differentiated judges just described.

In all cases, subjects responded using the digits 1 to 9, which were listed between the two adjectives defining each scale. There were also adverbial headings (e.g., *very*, *quite*, *slightly*). The left–right position of the desirable term was varied randomly across the 57 scales (and then reflected later where necessary so that higher ratings are always more favorable). In the internal data sets, each of the corresponding 114 single trait adjectives was rated for similarity of meaning on each of the 57 bipolar scales.

¹ The coefficient alpha reliability estimates for the two differentiated and the two broad composite judges are .77 and .73, respectively. In addition, we have replicated these findings with 47 subjects as judges of similarity of meaning of selected subsets of items. These subjects were partitioned, according to the number of neutral responses they made, as more differentiated or more broad. The more differentiated subjects again showed a greater tendency to separate Intellect and Emotional Stability variables from those related to the three large factors.

In the external data sets, subjects used the same scales to describe the target person "as accurately as possible."

One set of self-ratings was provided by 157 University of Oregon students who participated in this study in exchange for extra course credits. The other external data sets were all provided by 126 Swarthmore College students who were recruited by notices and paid. The Swarthmore students were asked to consider peers whom they knew well (preferably not from the college), including those whom they liked the most and those whom they liked the least.² They selected one person from each of these two groups, but not their best friend or worst enemy. They then rated three targets: themselves, the liked peer, and the not-liked peer.

Within each data set, the correlations among the 57 scales were analyzed by the method of principal components, followed by varimax rotation. In every case, we examined rotations of both five and six factors. We first considered the content of the factors themselves and then examined the size of the coefficients of congruence between corresponding factors in different data sets. For each of the external data sets, the five-factor solutions resembled the Big Five; moreover, the sixth factors were not consistent across the data sets. On the other hand, in the two internal data sets Emotional Stability was the sixth factor (smaller than Values) and was not included in five-factor solutions. Hence, we decided to retain five factors in the external data sets and six in the internal ones so as to be able to match five-factors consistently.

Results

The percentages of the total variance accounted for by each of the factors in each of the seven data sets are presented in Table 1. The differences in the total variances across all factors are large and systematic. The total variance is less for all of the four cases of restricted range than for any of the three cases of unrestricted range, and this difference is most pronounced for Factor II (Agreeableness). Of the two internal analyses, the values for the differentiated judges resemble those of the unrestricted external data set. In contrast, analyses of the broad judges include three large factors, with each of the smaller factors substantially reduced in size. The large variations in factor size necessarily involve some variation in their substantive nature. Nevertheless, there is substantial similarity in the factors common to the different data sets, especially for the three large factors.

To index the degree of similarity between the factors derived from each of the seven data sets, we used Tucker's coefficient of factor congruence (Harman, 1967, p. 257). Congruence varied substantially by factor, from a high of almost .90 for Factor III (Conscientiousness) to a low of approximately .60 for Factor IV (Emotional Stability). On the other hand, when averaged across the five matching factors in each data set, the mean congruence between pairs of data sets averaged about .75, whether derived from external data, from internal data, or from comparisons between the two.

For ease of presentation of the rather complex findings to be described in subsequent sections of this article, we used a standard or canonical representation, one that incorporates the prototypical core of the factor structures derived from each of the

Table 1
Seven Analyses of the Same 57 Scales: The Percentage of Total Variance for Each of the Factors

Data type and source	Factors					(Values)	Total
	II	III	I	V	IV		
Restricted range							
External							
<i>Self-descriptions</i>							
Oregon sample	11	9	10	6	6	—	42
Swarthmore sample	9	10	8	5	8	—	40
<i>Descriptions of others</i>							
Liked peers	13	13	9	7	6	—	49
Not-liked peers	11	15	10	7	5	—	48
Unrestricted range							
Liked and not-liked peers	23	14	8	8	4	—	58
Internal							
Differentiated judges	21	17	11	12	7	7	75
Broad judges	35	28	12	2	2	3	83

Note. All values are based on varimax-rotated principal components.

seven data sets. In aggregating across the seven data sets, however, we must consider both the factor variances (Table 1) and the inter-set congruence coefficients. Because the latter are so uniform, it is only the former that will have much influence on a composite representation. Moreover, although there are fewer internal data sets than external ones, the factor variances of the internal sets are larger; thus, a simple average of the seven sets will tend to approximate a composite structure that provides equal weighting of the two types of data sets (Gulliksen, 1950, pp. 312–327). These mean factor loadings are virtually identical to the pattern of loadings derived from the average (via Fisher's z transformation) of the correlations across the data sets; the congruence coefficients between the two types of composite representations are .99, .99, .99, .92, and .96 for Factors I through V, respectively. To save space, only the findings from the former analyses are described in this article.

The Three Large Factors

In this section, we examine the three large factors that were found in all seven data sets. Of the 57 scales, 40 generally had their largest loadings on these factors; we focus first on these 40 "primary" scales. The remaining 17 "secondary" scales, which are considered separately, load most consistently on the remaining factors (Intellect, Emotional Stability, and Values).

For each of the 40 primary scales and each of the three large factors, the average loadings across the seven analyses are listed in the Three large factors columns of Table 2. These factor loadings could conform to either of two different types of structural

² Liked least is a concession to the claim that many Americans, like Will Rogers, will not admit to disliking anyone. There seems to be considerable validity to this claim, as the grand mean for this target was 4.9 on a scale on which 5.0 is neutral. This target is referred to as *not liked* throughout this article.

patterns. The traditional conception of simple structure, embodied in the varimax rotation procedure, aims at isolating independent clusters of variables; in such a factor pattern, there are a few large loadings and many near-zero loadings. An alternative pattern is a circular one, with intermediate-sized loadings located between the factors. Despite our use of varimax rotations, it is clear that the loadings on the three large factors comes closer to a circular pattern. The highest loadings are surrounded in both directions by those that progressively decrease; moreover, such a progression applies to the same sequence of scales for each of the three factors. This suggests a circular type of relation among the scales associated with these factors. For any two factors, it would be easy to represent this in two dimensions. Is there any way to represent the circular relations for all three factors at the same time?

For this purpose, we followed a procedure used by Peabody (1984), who demonstrated that these factors could be deliberately rotated to a dimension of general evaluation (E) and two descriptive dimensions, called tight versus loose (T vs. L) and assertive versus unassertive (A vs. U). The estimated loadings on E, T versus L, and A versus U are listed in their respective columns of Table 2. The ratio of the loading on T versus L to that on A versus U was treated as the cotangent of an angle, with T versus L as the *x*-axis and A versus U as the *y*-axis. These angles (in degrees counterclockwise from T vs. L) are presented in the columns labeled Desirable pole and Undesirable pole.

In Figure 1, these angles are used to arrange the scales in two circles, an upper circle displaying their desirable poles and a lower circle their undesirable poles. Perpendicular to the two circles is the evaluation dimension. The bipolar scales all intersect at a neutral point halfway down, and the contrasting pole is on the opposite side of the other circle. Geometrically, the representation is a double cone—an inverted cone on top of an erect cone—meeting at the neutral point in the middle of the figure.

In Table 2 and Figure 1, the 40 scales are divided into six subsets of 6–7 scales each; these subsets generally correspond to those used in Peabody's (1987) a priori classifications. In most cases, the angular locations show clear gaps between the subsets. On the other hand, the conceptual distinction between the Impulse Control and Conscientiousness subsets is supported by the size of the evaluative component, which is consistently small for the Impulse Control scales and larger for the Conscientiousness scales. Two pairs of subsets are conceptual opposites (Assertiveness vs. Unassertiveness and Impulse Expression vs. Impulse Control); in these cases, the contrasting label could be applied to the opposite poles of the scales in the lower circle.³ The other two subsets, Affiliation and Conscientiousness, do not have such conceptual opposites, and their unfavorable poles are labeled simply *nonaffiliation* and *nonconscientiousness*.

Figure 1 displays the circular projections of each of the 40 scales onto the two descriptive dimensions. Because the scales are not equally distributed around the circle, one can infer the probable locations of each of the three rotated factors. Varimax factors tend to be aligned near relatively dense concentrations of variables. In Figure 1, the projections of the three factors should each be central to a sector of about 120°. These conditions are fulfilled by the relatively dense concentration in the lower right portion of the upper circle. The entire sector of 67°

between *discreet* and *hard-working* is bounded by clear gaps at both ends. Hence, the relevant factor, Factor III (Conscientiousness), should be consistently centered near the middle of this sector.

The pattern is different on the left-hand side of the upper circle, where there is a fairly continuous distribution of scales over a sector of 207° between *persistent* and *modest*. This sector can be summarized by two factors, but it is not clear where they will be located. The answer depends on the Impulse Expression scales, whose loadings differ between internal and external analyses. In internal analyses, the Impulse Expression scales have their largest loadings on Factor II (Agreeableness), along with Affiliation and Unassertiveness; only the Assertiveness scales have their largest loadings on Factor I (Surgency). In external analyses, the Impulse Expression scales and the Assertiveness scales have their largest loadings on a Factor I', which is shifted toward Impulse Expression; Factor II' shifts in turn toward Unassertiveness.

These differences in the factor locations can be traced back to the correlations between the Impulse Expression scales and the others. As an illustration, consider the most extreme case, Sociable–Unsociable. With internal data, this scale is almost identical to Warm–Cold (average $r = .84$), and its correlations decrease as one moves toward Assertiveness (average $r = .18$ with Bold–Timid). However, with external data (where all of the correlations tend to be much attenuated), Sociable–Unsociable is close to Talkative–Silent ($r = .46$), and it is as highly related to Bold–Timid ($r = .36$) as it is to Warm–Cold ($r = .31$).

Table 3 provides a direct comparison between the loadings from internal and external analyses. Note that the Impulse Expression scales are the only subset whose loading pattern shifts between the two types of analyses. Moreover, in both cases there were always some scales with high loadings on each of the first 2 factors and that are thus located between them; examples include Talkative–Silent and Spontaneous–Inhibited in the internal analyses and Lighthearted–Grim and Cheerful–Gloomy in the external analyses.

The Smaller Factors

In the previous section, we described the three large factors and their relations to the 40 primary scales. In this section, we consider the other factors and the 17 remaining secondary scales. The mean factor loadings for each of these secondary scales averaged across all seven data sets are presented in Table 4. Note that the loadings on Factor V (Intellect) and Factor IV (Emotional Stability) tend to fit the traditional pattern of simple structure; there are few intermediate loadings and thus there is not much to be expected from a circular representation of these two factors. Accordingly, we have not used the same kind of

³ Our labels are chosen to emphasize these relations. In Figure 1, the subsets of scales have capitalized labels corresponding to their desirable poles; uncapitalized labels are used for the undesirable poles. These labels cannot capture all of the specific implications of the separate desirable and undesirable versions; for example, Unassertiveness does not do justice to the features of nurturance or caring for others that is implied by some of the favorable terms (e.g., *kind*, *unselfish*).

Table 2
Factor Loadings and Angular Locations of the 40 Primary Scales Averaged Across the Seven Data Sets

Cluster and scale	Three large factors			Conceptual dimensions			Angular locations		Smaller factors	
	I	II	III	E	T vs. L	A vs. U	Desirable pole	Undesirable pole	V	IV
Assertiveness										
Persistent-Nonpersistent	42	05	37	42	14	35	67	247	12	-15
Ambitious-Unambitious	43	02	36	40	15	36	68	248	13	-26
Forceful-Submissive	60	-26	03	05	01	65	89	269	20	01
Self-confident-Unassured	53	08	25	40	01	43	89	269	07	28
Bold-Timid	73	-07	-08	17	-22	68	108	288	15	09
Active-Inactive	56	24	02	38	-27	39	125	305	13	-15
Impulse Expression										
Talkative-Silent	58	27	-21	26	-48	40	140	320	-06	-01
Spontaneous-Inhibited	56	20	-36	11	-54	42	142	322	12	15
(Frank-Secretive)	(33)	(32)	(-06)	(30)	(-32)	(15)	(155)	(335)	(-29)	(16)
Enthusiastic-Unenthusiastic	45	48	-07	45	-44	18	157	337	16	-03
Sociable-Unsociable	38	45	-03	43	-38	13	161	341	02	15
Cheerful-Gloomy	38	56	-03	51	-44	09	169	349	-02	27
Lighthearted-Grim	28	50	-24	29	-55	03	177	357	01	29
Affiliation										
Friendly-Unfriendly	20	67	06	57	-39	-13	198	18	07	17
Warm-Cold	16	71	-07	50	-50	-18	199	19	08	14
(Natural-Affected)	(10)	(39)	(01)	(31)	(-24)	(-09)	(200)	(20)	(13)	(19)
Generous-Stingy	09	63	-11	39	-46	-20	204	24	08	-03
Flexible-Inflexible	06	54	-11	32	-40	-19	205	25	24	13
Trustful-Distrustful	10	62	05	49	-35	-20	210	30	-04	17
Good-Natured-Irritable	04	65	13	54	-28	-26	223	43	00	29
Unassertiveness										
Lenient-Harsh	-19	62	-08	30	-36	-45	231	51	01	14
Kind-Unkind	-09	66	14	51	-25	-38	237	57	17	07
Unselfish-Selfish	-12	56	13	43	-19	-36	242	62	06	08
Cooperative-Uncooperative	-05	61	23	55	-15	-32	245	65	01	12
Polite-Rude	-13	56	32	54	-03	-37	266	86	06	-01
Peaceful-Quarrelsome	-22	47	24	40	-02	-42	267	87	06	35
Modest-Conceited	-38	37	19	24	04	-51	274	94	05	04
Impulse Control										
Discreet-Indiscreet	-35	16	42	24	33	-39	310	130	16	02
Cautious-Rash	-33	02	68	32	61	-32	333	153	-01	08
(Refined-Unrefined)	(-03)	(15)	(43)	(36)	(25)	(-10)	(339)	(159)	(19)	(-06)
Self-Controlled-Impulsive	-23	-05	72	34	65	-19	344	164	-05	09
Thrifty-Extravagant	-21	-11	53	18	53	-15	344	164	-06	06
Serious-Frivolous	-20	-15	62	22	61	-12	349	169	07	-12
Conscientiousness										
Responsible-Irresponsible	02	23	70	61	41	-10	347	167	10	00
Thorough-Careless	-04	04	76	49	58	-07	353	173	10	-03
Practical-Impractical	-06	00	74	44	59	-07	353	173	06	10
Orderly-Disorderly	02	04	71	49	52	-01	359	179	-05	-01
Organized-Disorganized	04	03	71	48	52	01	1	181	-07	-05
Logical-Illlogical	01	-05	64	37	52	02	2	182	21	04
Hard-Working-Lazy	19	19	54	54	26	08	17	197	05	-24

Note. E = evaluation, T vs. L = tight vs. loose, A vs. U = assertive vs. unassertive. The three scales in parentheses are not consistent across the data sets. To emphasize the circular relations, the larger factor loadings are listed in boldface. The angular locations are in degrees; all other values are factor loadings with decimals omitted.

transformation as in our presentation of the three large factors, but instead discuss the varimax-rotated factors directly.

Intellect. The broad judges show only the stub of an Intellect factor; instead, the relevant scales are associated with the three large factors. Some load on Factor III (Conscientiousness), as do the Impulse Control scales; such scales seem to reflect aspects of "controlled" intellect. Other scales load on Factor I (Surgency), on Factor II (Agreeableness), or on both, as do the Impulse Expression scales; these scales seem to reflect aspects of "expres-

sive" intellect. This distinction between controlled and expressive forms of judged intelligence is analogous to that between convergent and divergent forms of tested intelligence. In the remaining six data sets, the partition between scales reflecting each of the two types of intellect is substantially attenuated. Instead, the first 3 scales of both types, as listed in Table 4, generally have their largest loadings on a separate factor.

Each of the scales listed in Table 4 as being "related" has some distinctive feature. Among the scales related to controlled

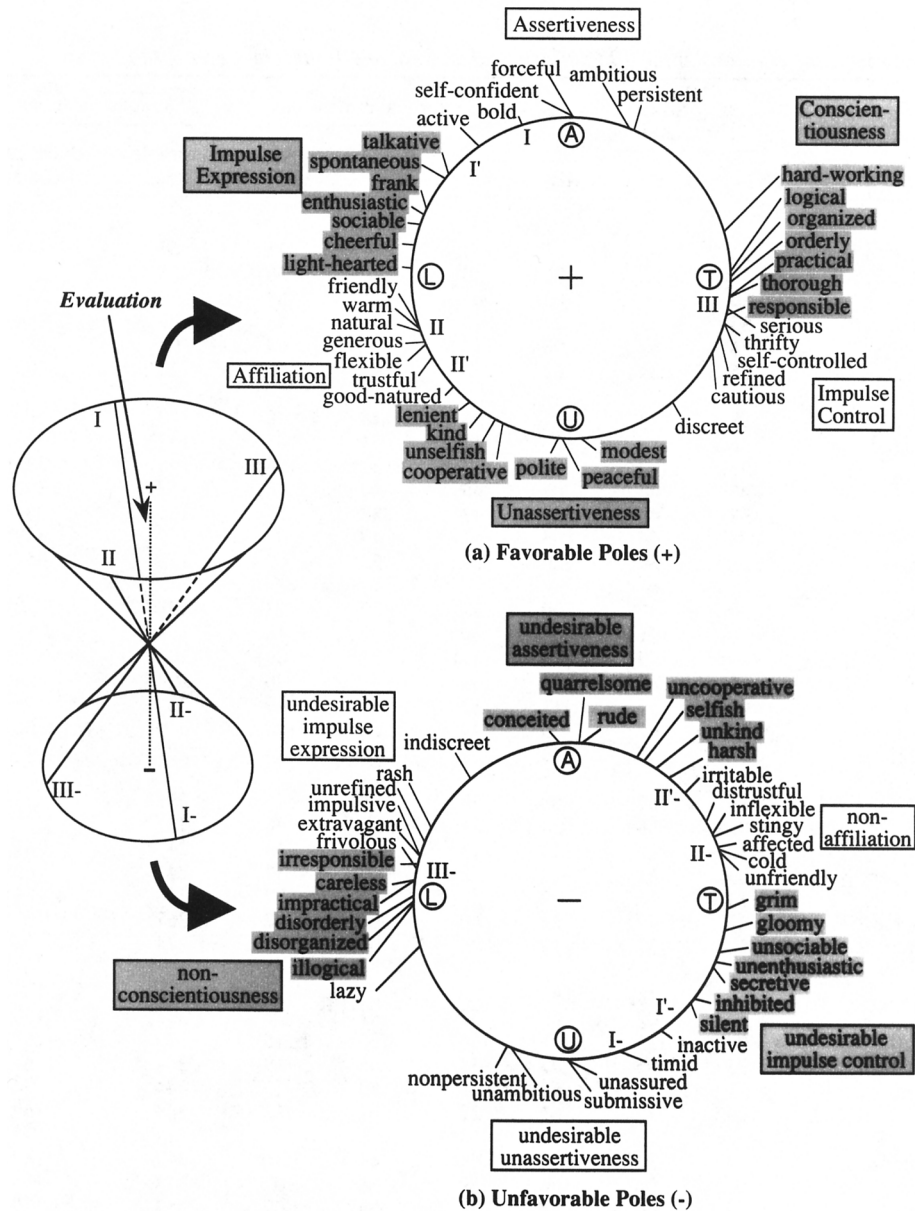


Figure 1. The circular projections of the 40 primary scales and of the three largest varimax factors onto the descriptive dimensions labeled tight (T) vs. loose (L) and assertive (A) vs. unassertive (U). (Shading or lack of shading is used to indicate the scales in each of the subsets.)

intellect, Capable-Incompetent sometimes has appreciable loadings on Factor I (Surgency), as well as on Factor III (Conscientiousness). The largest loadings for the scale Skeptical-Gullible are generally negative ones on Factor II (Agreeableness). The two related expressive intellect scales load inconsistently across the data sets.⁴

Emotional Stability. In data sets with a restricted range of target likability, Emotional Stability was nearly as large as the others; however, in sets with unrestricted range it was relatively small. In the selection of 57 scales, only 4 were considered a priori to be related to Emotional Stability, and thus these scales by themselves would only produce a small factor. When the

Emotional Stability factor becomes relatively large, it picks up variance from scales that are usually related to Factor II (Agreeableness). In cases of restricted range, there are always some such scales with appreciable loadings on Emotional Stability,

⁴ The scale Witty-Humorless has its largest loadings on Factor II (Agreeableness) with internal data, but on Factors V (Intellect) and I (Surgency) with external data. With the not-liked targets, this scale loads only on Factor I. The scale Independent-Conforming loads on Factor I in the two sets of self-ratings, but on Factor V in the three sets of peer ratings. Only in the two internal data sets did this scale have substantial loadings on both of these factors.

Table 3
The Six Subsets of Primary Scales: Mean Factor Loadings on Each of the Three Large Factors

Scale subset	Internal data			External data		
	I	II	III	I'	II'	III'
Assertiveness	63	12	22	51	-04	12
Impulse Expression	29	73	-16	47	26	-14
Affiliation	01	78	-04	14	53	01
Unassertiveness	-33	56	22	-10	55	15
Impulse Control	-30	-20	66	-19	08	53
Conscientiousness	02	-03	85	03	11	62

Note. To emphasize the circular relations, the larger factor loadings are in boldface. All values are factor loadings with decimals omitted.

but the particular scales are not consistent across the data sets. Indeed, many of the Agreeableness scales are fairly loosely related to each other, and an Emotional Stability factor could pick up some variance from several of them. However, with unrestricted range this does not occur.

In explanation, it can be shown mathematically that any restriction of range produces systematic effects on the means, variances, and correlations (e.g., Hills, 1957). Table 5 presents the means and variances for each subset of scales for each of the targets used in our external analyses. The differences between the liked and not-liked targets are listed in the last column of Table 5. The differences in means and variances are small for the Assertiveness, Impulse Expression, Impulse Control, and Conscientiousness subsets; large for Affiliation, Unassertiveness, and Values; and intermediate for Intelligence and Emotional Stability.

Thus, Table 5 shows differences that are of average size for the Emotional Stability scales, but quite large for the Factor II subsets, Affiliation and Unassertiveness. As a consequence, with unrestricted range the correlations become substantial among all these latter scales, including those that sometimes load on Factor IV. The result is a consolidation in Factor II of scales that were somewhat loosely related, and Factor IV is reduced to little more than the four a priori scales. At the same time, the sizes of their loadings on the consolidated Factor II are increased.

As an illustration of the effects of restriction of range on the correlations, we compared the average correlations for the separate liked and not-liked targets with those for the pooled liked and not-liked data set. The effect of unrestricted range in the pooled data is to increase the average correlations among the 14 Affiliation and Unassertiveness scales from .29 to .59, but among the 4 Emotional Stability scales only from .32 to .41. At the same time, the average correlation between the two sets of scales increases from .16 to .36. At the factor level, Emotional Stability (Factor IV) is not only reduced in size, but becomes less differentiated from Factor II (Agreeableness). That is, the Emotional Stability scales now load primarily on Factor II, and Factor IV becomes in effect an oblique satellite of Agreeableness.

Values. In internal analyses, there tends to be a Values factor that is small, but larger than Emotional Stability. Three scales (Honest-Dishonest, Moral-Immoral, and Fair-Unfair) have

their largest loadings on this factor. A similar factor also appeared in an earlier study of internal data (Peabody, 1984), where the second smallest rotated factor was defined by such scales as Idealistic-Opportunistic, Moral-Immoral, Honest-Dishonest, and Fair-Unfair. Thus, a small Values factor has now been replicated in internal analyses. On the other hand, a Values factor did not appear in the external analyses; instead, the relevant scales loaded on Factor II (Agreeableness) and Factor III (Conscientiousness).

Interim Summary

External versus internal judgments. The factors derived from external and internal judgments generally resemble each other. Nevertheless, there are several differences:

1. Internal representations account for higher percentages of the total variance, as would be predicted by any model of cognitive economy.
2. The broad judges show a tendency toward simplification, with the scales related to the small factors becoming relatively undifferentiated from the three large factors.
3. Internal representations include an additional Values factor.
4. In internal analyses, Factor I (Surgency) is limited to Assertiveness, and Factor II (Agreeableness) includes Impulse Expression as well as Affiliation and Unassertiveness. In external analyses, Factor I' shifts to include Impulse Expression as well

Table 4
Rotated Factor Loadings of Each of the 17 Secondary Scales

Scale	V	IV	II	III	I
Intellect					
Controlled aspects					
Central scales					
Intelligent-Unintelligent	60	-06	10	32	10
Reflective-Unreflective	49	-07	17	24	-09
Perceptive-Imperceptive	53	00	20	24	02
Related scales					
Capable-Incompetent	44	06	14	44	26
Skeptical-Gullible	38	03	-48	25	09
Expressive aspects					
Central scales					
Curious-Uninquisitive	57	-17	16	-03	27
Imaginative-Unimaginative	58	-07	21	-07	21
Broadminded-Narrowminded	51	03	38	-02	-03
Related scales					
Witty-Humorless	37	18	25	-10	26
Independent-Conforming	45	10	-02	-05	34
Emotional Stability					
Relaxed-Tense	-05	56	33	-06	09
Contented-Discontented	-11	45	36	29	17
Calm-Excitable	-10	48	32	40	-12
Stable-Unstable	-08	40	16	52	12
Values					
Honest-Dishonest (78)	13	07	40	30	-01
Moral-Immoral (75)	06	-04	35	37	-09
Fair-Unfair (55)	12	05	49	25	-07

Note. Values $\geq |.30|$ are in boldface. The values are factor loadings with decimals omitted, averaged across the seven data sets. The loadings for the Values factor (in parentheses) are averaged across the two internal data sets.

Table 5
Means and Variances of Each Subset of Scales

Scale	Number of scales	Self		Liked peer	Not-liked peer	Pooled: liked and not liked	Difference: Liked - not liked	
		Oregon	Swarthmore					
Rating means								
Primary								
Assertiveness	6	6.5	6.3	6.4	6.7	6.5	-0.3	
Impulse Expression	7	6.5	6.3	6.8	6.0	6.4	0.8	
Affiliation	7	6.8	6.6	7.2	4.1	5.7	3.1	
Unassertiveness	7	6.5	6.4	6.8	3.5	5.2	3.3	
Impulse Control	6	5.7	5.8	5.5	4.4	4.9	1.0	
Conscientiousness	7	6.7	6.4	6.2	5.0	5.6	1.2	
Secondary								
Intellect and related	10	7.1	7.2	6.9	5.0	6.0	1.9	
Emotional Stability	4	5.9	5.6	6.0	4.2	5.1	1.8	
Values	3	7.2	7.4	7.3	4.4	5.8	2.9	
Mean (57 scales)		6.6	6.5	6.6	4.9	5.7	1.7	
Rating variances								
							Within	Between
Primary								
Assertiveness		2.6	3.6	3.2	4.8	4.0	3.9	0.1
Impulse Expression		3.0	3.2	3.2	4.7	4.2	4.0	0.2
Affiliation		2.6	2.7	2.6	4.3	5.9	3.4	2.5
Unassertiveness		2.6	2.6	2.8	3.6	5.9	3.2	2.7
Impulse Control		3.2	3.5	3.7	5.3	4.8	4.5	0.4
Conscientiousness		3.0	3.7	4.6	5.6	5.6	5.1	0.5
Secondary								
Intellect and related		2.1	2.2	2.7	4.6	4.8	3.7	1.1
Emotional Stability		4.1	4.3	4.4	4.8	5.5	4.6	0.9
Values		1.9	2.5	2.9	4.8	6.0	3.8	2.2
Mean (57 scales)		2.7	3.1	3.3	4.7	5.1	4.0	1.1

Note. The response scale ranged from 1 to 9 and has been scored so that 9 is always the desirable pole of the scale. The variances of the pooled liked and not-liked targets have been partitioned into variance due to differences *between* the two targets and the average variance *within* each of the two separate targets.

as Assertiveness, and Factor II' shifts in turn toward Unassertiveness.

Restriction of range of target likability. Restricted range serves to decrease the size of the factors, especially that of Factor II (Agreeableness). Factor II scales become less cohesive, and some of them load on Factor IV (Emotional Stability). With unrestricted range, Factor II becomes larger and more consolidated. At the same time, Factor IV becomes smaller and less differentiated from Factor II.

Comparisons With Previous Findings

In the preceding sections, we examined the results from analyses of seven data sets all based on the same 57 scales. In this section, we consider the findings from the most relevant previous studies, all of which used other selections of variables. We restrict our analyses to studies that have used trait adjectives, in particular those that have included more than 20 variables and have made some attempt at a comprehensive coverage of the trait domain.

Studies Considered in This Section

Tupes and Christal. Tupes and Christal (1961) reported the findings from their analyses of eight data sets, including two by

Cattell, based on the 35 variables of Cattell (1947) or revisions thereof. These findings served to define the Big Five factors. The differences among these eight data sets are small compared with those between them and other findings. Accordingly, we simply consider the average results across their eight data sets.

Norman's 75 categories. Norman classified 1,431 trait adjectives into 75 categories. Goldberg (1982) included these terms in his inventory of 1,710 trait descriptors; the inventory was administered to 187 college students who rated the self-applicability of each term on an eight-step rating scale. Treating the categories as inventory scales, Goldberg analyzed their structure by using a wide variety of methods of factor extraction and rotation and found a near-invariant set of five factors.

Goldberg's 133 clusters. Focusing on a smaller set of more common trait adjectives, Goldberg classified 479 terms into 133 near-synonym clusters. Scale scores based on these 133 clusters are available from two samples of self ratings (Sample C and Sample D; $N_s = 314$ and 187, respectively), a sample of liked peers (Sample A; $N = 298$), and a data set that included descriptions of liked, neutral, and not-liked peers (Sample B; $N = 194$). The last sample provides another example of the effects of unrestricted range of target likability.

Peabody's trait-inference study. In earlier analyses of internal

data (Peabody, 1967, 1984), judges rated the likelihood of inferences from each of 120 single-trait adjectives to 55 bipolar scales. Each of the 6,600 items was rated by 10–20 judges. The initial selection of variables was aimed primarily at unconfounding evaluative and descriptive aspects of the traits and so emphasized Impulse Control versus Impulse Expression and Assertiveness versus Unassertiveness. To increase representativeness, other variables (e.g., for Affiliation and Conscientiousness) were added later.

Goldberg's 69 scales. Goldberg used a set of 69 bipolar scales (12–15 scales for each of the Big Five factors), both for analyses of self-ratings ($N = 95$) and for judgments of similarity of meaning. These analyses provide another comparison between the structures derived from external and internal data.

McCrae and Costa. McCrae and Costa (e.g., 1985b) have proposed a factor of Openness to Experience as Factor V in the Big Five structure. Initially, they used questionnaire statements, with those for Openness added to traditional ones for Neuroticism (Factor IV) and Extraversion (Factor I). In addition, they have used bipolar trait-adjective scales and added variables relating to Factors II and III (Agreeableness and Conscientiousness). Their studies are distinguished by samples that are large and that consist mainly of nonstudent adults. We considered McCrae and Costa's analyses of self ($N = 434$) and peer ratings ($N = 738$) based on 80 bipolar scales (McCrae & Costa, 1985c, 1987).

Factor Size

The studies we considered all include rotations of five factors, and their findings are reported in Table 6. An important preliminary finding is that the factors from each of these analyses could be identified as a version of the Big Five. For simplicity, the Values factor is not shown, because it appeared only in internal analyses. We focus first on the percentages of variance across all five factors, listed in the Total column of Table 6.

In general, the results in Table 6 from the previous data sets are consistent with our findings based on the 57 scales. Internal analyses always produce the largest percentages of total variance accounted for by the five factors. Restriction of range of target likability also has a substantial effect on the total variance; with restricted range the total is always smaller, whether for self-reports (averaging 41%) or for ratings of others (averaging 45%). The total variance is always larger with unrestricted range, averaging 60% for ratings of others. The relative predominance of the larger factors—particularly Factor II (Agreeableness)—over the smaller factors—particularly Factor IV (Emotional Stability)—is small in cases of restricted range, but becomes large with unrestricted range; this is especially true if one takes into account the overrepresentation of Emotional Stability variables in the Tupes and Christal (1961) studies. Indeed, the Tupes and Christal results stand out as the most discrepant from all the others in Table 6. We now try to show that these discrepancies can be attributed to the selection of the 35 variables by Cattell (1947).

Selection of Variables: The Cattell Set

The Tupes and Christal (1961) results were based on the 35 variables of Cattell (1947). For the majority of these variables,

it is possible to find clear counterparts in our own 57 scales, and these correspondences are presented in Table 7. For entries in the table that include more than one value, the smaller number stands for the cases of clear counterparts, and the larger number includes cases that are less clear cut.

The number of Cattell (1947) variables (35) is about 60% of the number of scales (57) used in our analyses. If the two selections were comparable, one would expect a similar ratio for each subset of scales. In representing Factors I and II (Surgey and Agreeableness), there is such a ratio, with 14–18 Cattell variables corresponding to 27 of our own. On the same basis, the Cattell variables should include about 7 for the Factor III subsets, about 6 for Intellect and related scales, about 1 for Culture, and about 2 for Emotional Stability. Instead, the second column of Table 7 shows that the Cattell variables include a drastic underrepresentation of Factor III and Intellect scales and an overrepresentation of scales related to Culture and Emotional Stability. These differences correspond precisely to the most discrepant results from the Tupes and Christal (1961) analyses shown in Table 6, where the percentages of variance are low for Factor III and high for Factor IV (Emotional Stability) and where Culture is substituted for Intellect as Factor V.

The Three Large Factors

We showed earlier that the three large factors could be rotated to evaluation and two descriptive dimensions, and that the variables associated with these three factors could then be projected onto the two descriptive dimensions in a circular representation. From an aesthetic standpoint, our circle might seem unsatisfying. The distribution of variables is uneven; in particular, there is a large gap at the upper right of the top circle in Figure 1. It would be easy to fill the gap by ignoring evaluation and using undesirable terms from the right-hand side of the lower circle.

Other circles. This was done in developing the best-known circle, that for interpersonal traits (Leary, 1957). Such traits were arranged in a circular representation summarized by the two dimensions of dominance–submission and love–hate. In concentrating on interpersonal traits and two factors, the implied hope was that these limitations would produce a compelling structure, and extensive efforts have gone into revisions of this circle (e.g., Kiesler, 1983; Wiggins, 1979, 1980). The Interpersonal Circle generally corresponds to Factors I and II (Surgey and Agreeableness), omitting Factor III (Conscientiousness) from our circle. The love half of the Interpersonal Circle corresponds to the desirable terms on the left-hand side of the upper circle in Figure 1; the hate half switches to their unfavorable opposites on the right-hand side of the lower circle.

For purposes of comparison, it would be useful to have other representations that might correspond to the two types of circles. The most influential attempt at a comprehensive coverage of personality characteristics is the classification of needs by Murray et al. (1938), and it has repeatedly served as a source of personality measures (Goldberg, 1971). Stern (1970) provided some evidence regarding the factor structure of these variables. Wiggins (1979, p. 410) rightly called this work “strangely neglected” and pointed out that Stern found a circular arrangement of the scales he developed to measure 30 of the Murray

Table 6
A Comparison of 17 Data Sets: Percentages of Variance for Each of the Five Factors

Data type and source	II	III	I	V	IV	Total	Number of scores combined
Restricted range							
External							
Self-descriptions							
Goldberg 133 clusters (Sample C)	8	7	8	5	5	33	1-9 ^a
Goldberg 133 clusters (Sample D)	9	8	9	5	6	36	1-9 ^a
McCrae & Costa 80 scales	8	11	8	7 ^b	6	40	
Swarthmore 57 scales	9	10	8	5	8	40	
Oregon 57 scales	11	9	10	6	6	42	
Goldberg 69 scales	8	7	10	7	10	42	
Norman 75 categories	14	10	13	7	10	54	1-51 ^a
Descriptions of others							
Goldberg 133 clusters (Sample A, liked peers)	12	9	9	5	5	40	1-9 ^a
McCrae & Costa 80 scales	12	11	7	7	7	44	
Swarthmore 57 scales (not-liked peers)	11	15	10	7	5	48	
Swarthmore 57 scales (liked peers)	13	13	9	7	6	49	
Unrestricted range							
Swarthmore 57 scales (liked & not liked)	23	14	8	8	4	58	
Goldberg 133 clusters (Sample B)	21	13	10	9	4	58	1-9 ^a
Tupes & Christal (average of 8 studies)	17	8	17	11 ^b	12	65	3-30 ^c
Internal							
Goldberg 69 scales	25	16	15	10	4	71	2 ^c
Swarthmore 57 scales (combined judges)	32	24	14	7	3	80 ^d	4 ^c
Peabody 55 scales	32	27	10	12 ^b	3	83 ^d	10-20 ^c

^a These variables were based on several items for each target.

^b Cases in which Factor V may be interpreted as something other than Intellect: Tupes and Christal (Culture); McCrae and Costa self-reports, and Peabody 55 Scales (Openness to Experience).

^c These variables were based on several judges for each item.

^d There was an additional Values factor for the Swarthmore combined judges (6%) and the Peabody 55 Scales (4%).

Table 7
Concordance Between Cattell's 35 Scales and Those in the 57-Scale Set

Present counterparts	Number of scales		Classification by Tupes & Christal (1961)					Scale no. from Cattell (1947)
	Present study	Cattell (1947)	I	II	III	IV	V	
I Assertiveness	6	4	3		1			4, 18, 9, 17
I/II Impulse Expression	7	3-5	3-5					16, 34, 5, 33, 28
II Affiliation	7	4		4				7, 23, 11, 12
II Unassertiveness	7	3-5		3-5				19, 1, 15, 22, 24
III Impulse Control	5	0						
III Conscientiousness	7	2			2			6, 20
IV Emotional Stability	4	6				6		8, 26, 13, 2, 30, 14
V Intellect								
Central scales	6	0-1					0-1	29
Related scales	4	1			1			25
Culture ^a	1	3					3	16, 31, 21
Values	3	1			1			27
Counterpart unclear	—	3	1	1			1	32, 3, 35
Total	57	35	9	10	5	6	5	

^a This is the scale Refined-Unrefined, which was classified in the Impulse Control subset because in our analyses it loaded primarily on Factor III.

needs. In contrast to the Interpersonal Circle, Stern included measures relevant to achievement and thus to Factor III (Conscientiousness). His analyses proceeded from 12 oblique factors to 3 second-order factors, of which the first—a general factor—was set aside. The other 2 factors were then used to plot a circular representation.

This last procedure resembles our own—very closely so, if one can assume that evaluation is included in Stern's general factor. Stern compared the Interpersonal Circle with his own (Stern, 1970, p. 63; reproduced in Wiggins, 1973, p. 486), and he expressed some optimism that the representations were similar. However, the similarity applies only to the parts of the circles that correspond to the left-hand side of the upper circle in our Figure 1. In the Interpersonal Circle, the other half includes traits describing an undesirable lack of Affiliation (hate), whereas in Stern's circle it includes desirable variables related to Factor III (Conscientiousness), including scales such as Achievement, Practicalness, Order, and Deliberation. Thus, Stern's circle resembles our own, not the Interpersonal Circle.

Factors I and II: Alternative versions. In most previous analyses of external judgments, the locations of Factor I (Surgency) and Factor II (Agreeableness) are located near the I' and II' positions in Figure 1. However, the locations of these factors can be shifted by the selection of variables. Suppose, for example, that one followed the policy of Norman (1963) and selected only variables that load most univocally on each factor, while discarding the rest. Using the findings presented in Table 2, one might select Warm–Cold and Flexible–Inflexible for Factor II (Agreeableness) and discard Sociable–Unsociable and Talkative–Silent, which are split between Factors I and II. Using the alternative I' and II' positions, however, Norman (1963) selected the counterparts of the latter two scales to represent Factor I' and discarded the former as being split between Factors I' and II'. In either case, one would not discover alternative locations for these factors, because variables located in the interstitial positions are no longer included in the variable set. (For example, using Norman's selection of variables, one could not show that these factors tend to be located near the I and II positions in internal analyses.)

Factor V

Among the five factors found in all of our data sets was one we interpreted as Intellect. Two alternatives now need to be considered, namely that Factor V should instead be interpreted as (a) Culture or (b) Openness to Experience.

Culture. The Culture interpretation was a historical accident. At the penultimate stage, Cattell (1946) still included variables related to rated intelligence, and these served to define the second largest factor. However, Cattell (1947)—following the debatable assumption that rated traits and tested traits are the same—then omitted these variables and substituted an actual intelligence test. When this test was omitted from the other studies considered by Tupes and Christal (1961), some of the remaining 35 variables provided the basis for a factor that they called Culture. Culture would seem to be a reasonable interpretation for two of these variables (cultured vs. boorish; polished vs. crude) and possibly for a third (aesthetically fastidious vs. lacking artistic feeling). A fourth variable (imaginative vs. logi-

cal) is different; it represents the contrast *within* intelligence between expressive and controlled aspects of that concept. When traits related to intelligence have been reintroduced in the other studies summarized in Table 6, Factor V is more strongly related to Intellect than to Culture. Indeed, the variables related to Culture generally load more highly on Factor III (Conscientiousness) than on Factor V, and hence one cannot assume that Intellect and Culture are the same factor. On the other hand, as demonstrated by McCrae and Costa (1985c), one cannot assume that rated and tested indexes of intelligence will define the same factor.

Openness to Experience. We now turn to the proposal by McCrae and Costa (1985b) that Factor V is Openness to Experience. What are the conditions necessary for interpreting a factor as Openness? Expressive intelligence seems to be a central part of the construct as defined by McCrae and Costa, who tend to select variables to represent expressive but not controlled aspects of intelligence. However, both types of variable must be adequately represented in order to test one condition for an Openness factor: These two types of variables should have their largest loadings on separate factors and not on the same (Intellect) factor. This condition was not met in most of the studies summarized in Table 6, where in 14 out of 16 relevant cases the largest loadings of the variables relating to both controlled and expressive aspects of intelligence tended to be on the same factor. A case in point is the peer-rating study of McCrae and Costa (1987), where scales for openness and for controlled intelligence loaded on the same factor.

Findings that do meet this condition are of two kinds: (a) some analyses of internal data and (b) external data based on older adults. In both cases, the key may be that the scales related to controlled intelligence load primarily on Factor III (Conscientiousness). In McCrae and Costa's (1985c) analyses of self-ratings, most of the openness variables loaded primarily on Factor V, whereas the scales measuring controlled intelligence generally had smaller loadings on this factor than on Factor III. Thus, only one of McCrae and Costa's two studies meets this condition for an Openness factor.⁵

Factor IV (Emotional Stability)

In the English lexicon, relatively few trait adjectives are directly related to Emotional Stability—indeed, only 4 of the 57 scales selected as representative of common trait terms. These 4 a priori Emotional Stability scales by themselves would produce only a small factor. On the other hand, many of the previous studies included a larger representation of Emotional Stability scales. However, these scales are not always defined by common trait adjectives; thus, the rationale for representing personality characteristics in this way is abandoned. For example, the 35 Cattell variables include 6 for Emotional Stability;

⁵ Perhaps openness is more applicable to the older adults studied by McCrae and Costa (1985b) than to college students, who may not differ as much in this respect. Specifically, the watershed changes in our society in the late 1960s may have served to separate the generation who were already adults into those more and those less open to these changes in a way that would not present the same challenge to the students who came afterwards.

however, these include neurotic versus not so and hypochondriacal versus not so, which are hardly common trait adjectives.

The findings in Table 6 from previous studies with restricted range are generally consistent with our own; the Emotional Stability factor is nearly as large as the other four. The relatively smaller size of Factor IV in previous studies with unrestricted range is also consistent with our own findings if one takes into account the overrepresentation of Emotional Stability scales in the Tupes and Christal (1961) studies. To ensure that the Emotional Stability factor would be large in analyses based on unrestricted range, one would have to include more variables, in which case the selection of variables would have to be based on some rationale other than that of trait representativeness. For example, one might include terms describing temporary states.

Some Implications of Our Findings

Utility of a Standard Representation

A comprehensive structure of personality characteristics provides an appropriate context for the interpretation of any more limited subset. The simplest case involves the interpretation of a single factor, such as Factor V. This factor may be interpreted as Culture if no variables relating to either expressive or controlled aspects of intelligence are represented. Alternatively, the factor may be interpreted as Openness if the variables relating to intelligence tap primarily the expressive aspects of that concept. However, if variables related to both aspects are included, it is harder to escape an interpretation of the factor as Intellect or its equivalent.

More complex cases involve two or more factors. We have discussed the Interpersonal Circle, which is based on Factors I and II (Surgency and Agreeableness) and therefore omits Factors III, IV, and V (Conscientiousness, Emotional Stability, and Intellect). Another example is the factor representation popularized by Eysenck, which initially included only Neuroticism and Extraversion (e.g., Eysenck & Eysenck, 1969). Neuroticism corresponds to Emotional Stability (Factor IV), labeled in the reverse direction. Extraversion *could* correspond to a descriptive dimension unconfounded with evaluation, running between the upper left (Impulse Expression) and the lower right (Impulse Control) in either or both of the circles in Figure 1. Typically, however, the factor is confounded with evaluation, contrasting favorable Extraversion with unfavorable Introversion, and so is another label for Factor I'. In more recent versions of Eysenck's model, Extraversion and Neuroticism have been joined by a third factor, called Psychoticism (e.g., Eysenck, 1978). Psychoticism could correspond to Factors II, III, or V or to any combination of these factors. Indeed, one recent analysis of this factor in the context of the Big Five representation suggests that it combines the undesirable poles of Factors II and III (McCrae & Costa, 1985a).

Recently, another use of Factors I' and IV has been introduced by Tellegen and his colleagues for the domain of mood-related terms (Tellegen, 1985; Watson & Tellegen, 1985). Tellegen proposed that the relations among these terms can be summarized by two factors, which he calls Positive Emotionality and Negative Emotionality. Positive Emotionality contrasts such terms as *active, energetic, enthusiastic, and alert* (Factor

I'+) with those such as *quiet, sluggish, dull, and glum* (Factor I'-). Negative Emotionality contrasts such unstable states as *distressed, fearful, jittery, and nervous* (Factor IV-) with stable states such as *calm, placid, and relaxed* (Factor IV+). From our perspective, the two factors that summarize the relations among mood terms seem to be easily assimilated within the five-factor structure of traits.

The advantages of some standard or canonical representation should now be apparent: It provides a context for the integration of findings from studies that include more limited sets of variables. Moreover, by representing a comprehensive collection of variables in terms of their conceptual similarities and differences, it permits investigators to accrue both convergent and discriminative evidence for the validation of their measures. Given the desirability of some structural representation, what are the relative advantages of a particular type?

Role of Evaluation in Personality Factors

All of the Big Five factors are confounded with evaluation. In the structural model displayed in Figure 1, the three largest of the varimax factors have been transformed into a purely evaluative dimension and two descriptive dimensions. Each of these two types of representations can be expected to have some advantages. We can see two advantages for the evaluation-transformed factors, one practical and the other theoretical. The practical case rests on the usefulness of expressing the relations among three factors within two dimensions. We have shown that the variables associated with the three largest varimax factors are systematically related and that these relations can be captured in a circular representation. However, it is difficult to display a pattern of relations involving three factors in the same figure. Yet when evaluation is set aside as a general dimension, the remaining relations among the three factors can be conveniently represented in two descriptive dimensions.

At the theoretical level, the explicit separation of evaluation from the descriptive factors provides a conceptual clarity that is missing in the varimax factors, which confound evaluative and descriptive aspects. Indeed, when two or more varimax factors form the basis of a representation, the conceptual ambiguity may be compounded; for example, because the love-hate dimension of the Interpersonal Circle is so highly related to evaluation, the circle is divided quite sharply into favorable and unfavorable halves. Moreover, by making evaluation explicit it becomes possible to test those theories that emphasize the importance of evaluation in the human judgment process, for example, theories of person perception, which are sometimes formulated exclusively in terms of evaluation.

Similarities and Differences in Factor Structures

The analyses presented in this article illustrate an application of the experimental method to an arena in which it has not traditionally been used. Factor structures have generally been interpreted rather uncritically. In comparing the findings from different analyses, it is often assumed that factors given similar labels are similar and that factors given different labels are different. Indeed, in the literature differences in factor size and content are rarely mentioned, much less emphasized, and to

our knowledge no one has studied the variables that influence them. In contrast, we were able to account for personality factors on the basis of such influences as external versus internal judgments and restricted versus unrestricted range of target likability.

In addition to the independent variables we have studied, there are others that may also affect the size or content of personality factors. For example, in our external analyses we have not studied the effects of such variables as length of acquaintance, closeness of relationship, or role in the relationship. Moreover, in our analyses of internal data we have not distinguished between judgments of semantic similarity or trait inferences. Furthermore, we did not control for differences in methods of factor extraction (e.g., component vs. factor models) or in the particular rotational procedures that were used.⁶ Although we doubt that the effects of such procedures will turn out to be as strong or as pervasive as those of the influences we have analyzed here, analyses of such variables provide an agenda for future studies of this type.

Overview and Conclusions

We began this article with a brief history of the lexical approach to personality-trait structure, including the rationale for analyzing the natural language of personality. We described the seminal contributions of Allport and Odbert (1936), Cattell (1947), Tupes and Christal (1961), and Norman (1963) leading to what has now come to be called the Big Five representation of personality traits. Although there has been a gradual convergence in views among investigators of personality structure about the *number* of orthogonal factors necessary to account for the interrelations among most English trait-descriptive terms, there appears to be less agreement about the exact nature of these five broad domains. Some of the apparent disagreement may be merely terminological, in which case one should expect increased consensus as investigators converge on a common scientific vocabulary. On the other hand, some of the disagreement may arise from variations in the factor structures that have been obtained in different investigations.

The major purpose of the present article was to begin the scientific analysis of these differences so as to understand why they occur, thus enabling investigators to predict or control their occurrence in future studies. Metaphorically, one can view the Big Five representation, like a piece of classical music, as having a theme and variations on that theme. We seek to understand the principles underlying the variations that are found in the scientific literature on structural representations of personality traits.

In the present article, we have focused on three major influences on factor structures; we viewed them as three independent variables in this new type of experimental psychometrics. We make the implicit assumption that these are *among* the most powerful influences, if not *the* most powerful ones. This assumption is based on years of study of these and other potential influences, ones that have not turned out to produce effects as strong or as consistent as the three under study here.

Our first independent variable is quite generally recognized as being important in determining the size and nature of factors, namely the selection of variables. Without enough vari-

ables with similar features included in the analysis, no factor reflecting their covariations can be obtained. In addition, when only a narrow range of variables are included as representatives of a broad domain, the factor representing those variables will inevitably reflect their peculiar features; moreover, because of the constraints of orthogonality imposed by typical rotational criteria, the locations of other factors can change as well. To control for this independent variable in our own analyses, we always included the same set of 57 bipolar scales, a set designed to be at least approximately representative of common English trait-descriptive adjectives.

Moreover, when we compared our findings with those from previous studies, we analyzed the correspondence between the variables that had been previously selected and those in our own representative set. We were able to show that the overrepresentation of variables associated with psychopathology in the set of variables devised by Cattell (1947) and later analyzed by Tupes and Christal (1961), Norman (1963), and others led to the increased variance of Factor IV (Emotional Stability) in their studies as compared with our own. Of more theoretical importance, we showed how the selection of variables can determine the interpretation of Factor V, which we have labeled *Intellect*: If all variables representing both expressive and controlled aspects of intelligence are excluded from the analyses (as they were in the Cattell set), then one might obtain a factor that can be labeled *Culture*. On the other hand, if primarily expressive aspects of intelligence are included, then one might obtain a factor that can be labeled *Openness*.

Whereas the selection of variables has long been recognized as being important in determining the size and location of factors, the effects of our two other independent variables have received relatively little attention. The first of these involves the kind of data used to assess covariations among the variables under study. If these relations are assessed through the judgments of native speakers rating the semantic similarity among the trait terms, the resulting factor structures are called *internal* representations. On the other hand, if the relations are assessed through the intercorrelations of the terms when they are used to describe oneself or others, the resulting structures are called *external* representations.

Contrary to the implicit assumption of the systematic distortion hypothesis, we have found some important differences in the factor structures derived from internal and external judgments. The total variance associated with factors derived from

⁶ In all of our analyses, we have compared the structures produced by different procedures for factor extraction and rotation, and we typically compared rotations of five, six, and seven factors. For example, in our analyses of Norman's 75 categories we systematically compared five methods of factor extraction (principal-components, principal-factors, alpha, image, and maximum-likelihood factoring), each rotated orthogonally (varimax) and obliquely (oblimin). When averaged across the corresponding factors in each pair of analyses, the mean correlation of the factor scores between the orthogonal and oblique rotational procedures (holding constant the method of factor extraction) ranged from .991 to .995, and the mean intercorrelations among the five methods of extraction (holding constant the procedure for rotation) ranged from .950 to .996. In all of our analyses, we have never discovered any influences on factor representations that produce effects as substantial or as consistent as those reported in this article.

internal data is much greater than that derived from external data. Whereas each of the Big Five factors derived from external analyses of a representative selection of variables tend to be roughly equal in size, those derived from internal data include three quite large factors and three of much smaller size. (The sixth factor found solely in internal data includes variables related to Values, such as honest-dishonest.) In general, then, internal representations are both more simple and more schematic than those derived from external data. We explain these differences on the basis of the cognitive economy associated with judgments of semantic or conceptual relations.

As our final independent variable, we studied the degree to which the targets in external analyses are similarly evaluated by the judges. When the range of target liking is restricted to, say, descriptions of oneself or one's friends, then each of the resulting factors tends to be of equal size, and the factors tend to be clearly differentiated from each other. In analyses of evaluatively heterogeneous targets, the factors change their size (and to some extent their locations) as a function of their evaluative implications. For example, when both liked and disliked targets are analyzed together, the variance associated with Factor II (Agreeableness) tends to be much larger than in analyses of either kind of target alone.

Whereas it is relatively easy to convey information about the size of factors, it is considerably more difficult to describe their content, unless the reader can see their locations in multidimensional space. If that space is restricted to only two dimensions, the relations among variables are relatively easily displayed in a figure, and therein lies one of the major attractions of circumplex models. However, we have consistently found that the variables associated with Factor I (Surgency), Factor II (Agreeableness), and Factor III (Conscientiousness) are distributed rather uniformly throughout that three-dimensional space, rather than being concentrated in tight clusters. Unfortunately, it is difficult to display such a pattern in a single figure.

On the other hand, these three factors can be deliberately rotated to general Evaluation and two descriptive dimensions (assertiveness and tight vs. loose impulse expressiveness), after which the two descriptive dimensions can easily be displayed in two-dimensional space. We make much use of this model, both in displaying the findings from our own analyses and in integrating the findings from previous investigations. Moreover, we argue that this factor structure may be particularly useful in integrating findings from studies of person perception, many of which are analyzed solely in terms of general Evaluation. In any case, we argue for the necessity of *some* standard structural representation, so as to be able to specify the ways that one pattern of findings is similar to and differs from another pattern. Only through analyses of such themes and variations within a common framework will we achieve our taxonomic goal, namely the development of a compelling structure of personality traits.

Summary

We investigated three determinants of the factor structures of personality traits. The most important is the selection of variables, which was controlled by including the same set of 57 representative variables in analyses of seven data sets. Variants of

the Big Five factors were always found. The factor similarities were particularly strong for the three largest factors, traditionally called Agreeableness (Factor II), Conscientiousness (Factor III), and Surgency (Factor I). These three factors were transformed into general evaluation and two descriptive dimensions, assertive versus unassertive and tight versus loose; these latter dimensions were used to show circular relations among the variables. A second determinant involves the use of external judgments about actual people or internal judgments about the conceptual relations among traits. The factor structures based on the two types of judgments are generally similar, but those based on internal judgments tend to be simpler. For example, with internal judgments the smaller factors—such as Intellect (Factor V) and Emotional Stability (Factor IV)—are not so clearly differentiated from the three large factors. The third determinant involves the degree of restriction of the sample to evaluatively homogeneous targets, such as oneself and one's friends. Such restriction of range reduces the size of all factors, but especially that of Factor II (Agreeableness). The findings were compared with those from previous studies, and it was shown that those findings are easily integrated within this framework.

References

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1, Whole No. 211).
- Austin, J. L. (1957). A plea for excuses. *Proceedings of the Aristotelian Society*, 57, 1-30.
- Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, 38, 476-506.
- Cattell, R. B. (1946). *The description and measurement of personality*. New York: World Book.
- Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika*, 12, 197-220.
- Digman, J. M., & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology*, 50, 116-123.
- Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research*, 16, 149-170.
- Eysenck, H. J. (1978). Superfactors P, E, and N in a comprehensive factor space. *Multivariate Behavioral Research*, 13, 475-482.
- Eysenck, H. J., & Eysenck, S. B. G. (1969). *Personality structure and measurement*. San Diego, CA: Knapp.
- Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, 44, 329-344.
- Goldberg, L. R. (1971). A historical survey of personality scales and inventories. In P. McReynolds (Ed.), *Advances in psychological assessment: Vol. 2* (pp. 293-336). Palo Alto, CA: Science and Behavior Books.
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141-165). Beverly Hills, CA: Sage.
- Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment* (Vol. 1, pp. 203-234). Hillsdale, NJ: Erlbaum.

- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Harman, H. H. (1967). *Modern factor analysis* (2nd ed.). Chicago: University of Chicago Press.
- Hills, J. R. (1957). Within-groups correlations and their correction for attenuation. *Psychological Bulletin*, *54*, 131–133.
- John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, *2*, 171–203.
- John, O. P., Goldberg, L. R., & Angleitner, A. (1984). Better than the alphabet: Taxonomies of personality-descriptive terms in English, Dutch, and German. In H. Bonarius, G. Van Heck, & N. Smid (Eds.), *Personality psychology in Europe: Theoretical and empirical developments* (pp. 83–100). Lisse, The Netherlands: Swets & Zeitlinger.
- Kiesler, D. J. (1983). The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. *Psychological Review*, *90*, 185–214.
- Klages, L. (1932). *The science of character* (W. H. Johnston, Trans.). London: Allen & Unwin. (Original work published 1926)
- Leary, T. F. (1957). *Interpersonal diagnosis of personality: A functional theory and methodology for personality evaluation*. New York: Ronald Press.
- McCrae, R. R., & Costa, P. T., Jr. (1985a). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, *6*, 587–597.
- McCrae, R. R., & Costa, P. T., Jr. (1985b). Openness to experience. In R. Hogan & W. H. Jones (Eds.), *Perspectives in personality* (Vol. 1, pp. 145–172). Greenwich, CT: JAI Press.
- McCrae, R. R., & Costa, P. T., Jr. (1985c). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*, *49*, 710–721.
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81–90.
- Murray, H. A., et al. (1938). *Explorations in personality*. New York: Oxford University Press.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*, 574–583.
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology Monographs*, *7*(4, Whole No. 644).
- Peabody, D. (1984). Personality dimensions through trait inferences. *Journal of Personality and Social Psychology*, *46*, 384–403.
- Peabody, D. (1987). Selecting representative trait adjectives. *Journal of Personality and Social Psychology*, *52*, 59–71.
- Shweder, R. A., & D'Andrade, R. G. (1980). The systematic distortion hypothesis. In R. A. Shweder (Ed.), *New directions for methodology of social and behavioral science: Vol 4. Fallible judgment in behavioral research*. San Francisco: Jossey-Bass.
- Stern, G. G. (1970). *People in context: Measuring person-environment congruence in education and industry*. New York: Wiley.
- Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–716). Hillsdale, NJ: Erlbaum.
- Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (Technical Report ASD-TR-61-97). Lackland Air Force Base, TX: U.S. Air Force.
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*, 219–235.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
- Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*, 395–412.
- Wiggins, J. S. (1980). Circumplex models of interpersonal behavior. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 1, pp. 265–294). Beverly Hills, CA: Sage.

Received June 21, 1988

Revision received February 16, 1989

Accepted February 17, 1989 ■