Journal of Personality and Social Psychology
1999, Vol. 76, No. 4, 613–627

0022-3514/99/$3.00

Hierarchical Subcomponents of the Big Five Personality Factors: A Cross-Language Replication

Gerard Saucier
University of Oregon

Fritz Ostendorf
Universität Bielefeld

An ideal structural representation of personality attributes would include more than just broad-bandwidth factors. Specific subcomponents help define broad factors while enhancing the fidelity of the representation. There has been no consensus with regard to the necessary specific subcomponents of the Big Five. This problem was addressed by analyzing 2 representative lexical data sets, one involving English adjectives and the other involving German adjectives. Large samples (*N*s of 636 and 775) were used in classifying a selection of 500 adjectives in each language by Big Five domains, and within each domain and language, the terms were factor analyzed with promax rotation. Ratings by 22 bilinguals of correspondence between the adjectives in English and German factors indicated 18 distinct content themes common to personality description in the 2 languages. The 18 subcomponents delineate necessary features of a more finely faceted measurement model for the lexical Big Five factors.

What are the most important ways in which people vary? This focal question for personality psychology is more complex than it might first appear. Which of the "ways" are intended: observable attributes—the phenotype—or instead variation in dynamic internal "personality processes"? If one limits oneself to the phenotype, there is still great potential for complexity. Does one mean people's temporary conditions, their social roles and evaluations by others, and their appearance and physical characteristics, as well as their temperament, character traits, and abilities (Saucier, 1997)? This article focuses on yet another major source of complexity: The ways in which people vary might be viewed from the perspective of a few independent, broad-bandwidth constructs or from the perspective of a larger number of more specific but not entirely independent constructs: lower level oblique factors. What perspective is taken on these matters affects, to some degree, the answers provided by the tool of factor analysis.

Many researchers have adopted the view that *phenotypic* personality variation is meaningfully distinguished in terms of five broad orthogonal factors labeled the Big Five (Digman, 1990; Goldberg, 1990, 1993b; John, 1990; Ostendorf, 1990a; cf. McCrae & Costa, 1987). Some or all of these five factors—Extraversion (Factor I), Agreeableness (Factor II), Conscientiousness (Factor III), Emotional Stability (Factor IV), and Intellect–Imagination (Factor V)—have emerged independently in "lexical" studies of the factor structure of adjectival disposition descriptors within a number of languages, including English (Goldberg, 1990; Saucier & Goldberg, 1996a), German (Ostendorf, 1990a), Dutch (De Raad, Hendriks, & Hofstee, 1992), Czech (Hrebickova, Ostendorf, & Angleitner, 1995), Polish (Szarota, 1996), and Russian (Shmelyov & Pokhil'ko, 1993). Moreover, studies in additional languages, such as Spanish (Benet-Martinez & Waller, 1997), Hungarian (Szirmak & De Raad, 1994), and Filipino (Church, Reyes, Katigbak, & Grimm, 1997), have produced results with high comparability to the Big Five, especially with respect to the first three factors (Extraversion, Agreeableness, and Conscientiousness; Peabody & Goldberg, 1989; Saucier, 1997).

The broad Big Five factors provide parsimony and fulfill an integrative function. Any more specific personality construct is likely to correlate substantially with at least one of the Big Five and can be grouped with other constructs that also relate most highly to the same Big Five factor. On the other hand, broad factors have their limitations. They are composed of many variables and can possess definitional ambiguity. As noted by Block (1995) and John (1990), observers differ in the psychological meaning they give to each of the Big Five factors. For example, Extraversion can be thought of as a composite of sociability and assertiveness (as well as other related constructs), but some (e.g., Costa & McCrae, 1992) see sociability as most central, whereas others (e.g., Goldberg, 1993b; Peabody, 1987) see assertiveness as most central. Although the factors are usually labeled with a single term, plumbing the psychological meaning of one of the Big Five is a cognitive task of considerable complexity. A broad factor is not so much one thing as a collection of many things that have something in common. It is easy to be overly schematic, ignoring the diverse character of the variables contained within a broad factor. A better way to understand each factor might be to characterize its crucial subcomponents, which, although correlated, are conceptually distinct.

Gerard Saucier, Department of Psychology, University of Oregon; Fritz Ostendorf, Department of Psychology, Universität Bielefeld, Bielefeld, Germany.

This work was supported by Grant MH-49227 from the National Institute of Mental Health. Lewis R. Goldberg provided the American data set for Study 1, along with valuable comments on early drafts. Delia Lang assisted in translation of the materials used in Study 2, and Erika Westling assisted in the recruitment of bilingual participants for Study 2.

Correspondence concerning this article should be addressed to Gerard Saucier, Department of Psychology, 1227 University of Oregon, Eugene, Oregon 97403-1227. Electronic mail may be sent to gsaucier@oregon.uoregon.edu.

In this way, identification of specific subcomponents can help to clarify conceptualization of the broader factors (Briggs, 1989). Because broader factors blend together subcomponents that might be distinguished from one another, some of the finer features of personality description are lost when one is making only a few broad distinctions. Such finer features might reflect genetic sources of variation beyond those bearing on broad-level factors (Jang, McCrae, Angleitner, Riemann, & Livesley, 1998). A representation of personality structure that makes the finer features explicit potentially offers higher fidelity in personality description. When categories are narrower, the exemplars for each are more similar, enhancing diagnostic value for specific instances; the disadvantage of such high-fidelity representation is that many subordinate categories are needed to represent the domain, decreasing efficiency (John, Hampson, & Goldberg, 1991). Broad-bandwidth constructs, on the other hand, sacrifice fidelity to gain efficiency. A representation combining broad and narrow constructs may be an optimal compromise: The broad-bandwidth level offers higher efficiency (i.e., parsimony), whereas the narrow level offers higher fidelity.

One example of the gain in fidelity involves the meaning of those scores that fall in the middle of the distribution on a measure of a broad (e.g., Big Five) factor. Whereas the extremes of the distribution would typically include individuals who are positioned uniformly high or low on all of the subcomponents of that factor, the middle of the distribution would include a mélange of (a) individuals scoring midrange on all of the subcomponents and (b) individuals scoring high on some subcomponents but low on others. To the extent that subcomponents are measured reliably, measuring subcomponents affords valuable information about middle scorers.

A possible collateral benefit of measuring subcomponents pertains to predictive validity. As diverse commentators (e.g., Goldberg, 1993a; Hough & Schneider, in press; Mershon & Gorsuch, 1988) have pointed out, the amalgamation of measures into broad factors leads to a loss of specific variance, possibly lowering the overall validity of the composite. In other words, if one seeks optimal prediction, one should use as many specific sources of variance as statistical power—and thus sample size—will permit (Goldberg, 1993a). The quest for optimal prediction will eventually lead to levels far more specific than a few subcomponents of the Big Five; a few aggregated subcomponents are likely to predict more powerfully than the single broad factor into which they are agglomerated.[1] Even in cases in which more specific variables provide little predictive gain over the common factors, it can be useful to know which aspect of the common factor is responsible for the bulk of the correlation, thus permitting more conceptual clarity. Similar principles would apply to the diagnosis of personality disorders; there are good arguments in favor of using a large number of specific constructs rather than just a few broad ones (e.g., Widiger, Trull, Clarkin, Sanderson, & Costa, 1994).

What, then, are the necessary subcomponents of each of the Big Five? The present article documents one attempt to answer this question. First, we established a basis for discovering meaningful personality factors by defining relatively representative sets of personality adjectives. Then, beginning with factors whose robustness is well established (the Big Five), we identified specific subcomponents of the broad factors. To ensure the robustness of our findings, we conducted our analyses not only in two different samples but in two different nations and two different languages.

Conceptualizations of Big Five Subcomponents

NEO Personality Inventory Facets

The best-known model of specific subcomponents of the broad five factors is that represented by the "facet" scales of the revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992). For each of the five broad "domains" (factors) on the inventory, six facet scales were developed to measure specific traits included in the domains; the specific traits chosen were based on reviews of the literature (Costa, McCrae, & Dye, 1991; McCrae & Costa, 1983). Although the choice of facets was reasonable, the authors conceded that "the facets proposed are not the only possible ones" (Costa et al., 1991, p. 888). Indeed, it would be quite astonishing if there were exactly six facets for each of the broad personality factors. The structure of personality attributes doubtless reflects some combination of sociocultural construction and biological–evolutionary influence, and, in either case, nature—even human nature—is not typically so neat.

By what evidence would the number and content of such subcomponents be checked and revised? One might argue for revision on the basis of the research literature, which gradually changes over time. But the literature often remains anchored in those emphases it acquired early in its development. Indeed, personality inventories in common use also tend to generate anchoring effects on the discipline. Researchers and practitioners become accustomed to conceptualizing along the lines set out by the model embodied in a particular inventory, with all of that model's implicit assumptions; a self-reinforcing circle develops supporting the use of the privileged constructs while ignoring those constructs whose content lies outside the model.

Lexical Criteria

One important source of evidence against which to assess any theorized classification of personality attributes into subcomponents is the natural language (Norman, 1967; Tellegen, 1993). Personality attributes are observable and socially meaningful phenomena, so high overlap between lay and expert descriptive models is to be expected. However, there is a crucial difference between these two types of models. Expert models rely, for better or worse, on the expert's subjective insights and predilections. But in the development of a lexical model, item and variable selection can be taken out of the researcher's hands and delegated to disinterested but authoritative parties (e.g., dictionaries or reliable aggregate judgments of laypersons). Thus, although lexical models

[1] This tendency generates a paradox: That a specific facet is predictive suggests that it is valid. Yet, a common theme in factor-analytic studies has been that factors beyond the first five are not replicable (i.e., not reliable across samples). How can a facet be valid when it cannot be shown to be reliable? The paradox might be resolved by showing that the facets have some reliability, even if they are not reliable enough to meet factor-analytic replication criteria (e.g., that of Everett, 1983), or by showing that facets are sufficiently replicable when derived by an alternative method, such as that used in the present study.

may be subject to their own forms of bias, they can provide a useful check on the biases of expert models. Examples are provided by the constructs of Agreeableness and Conscientiousness: Before the corrective influence of the lexical Big Five, which, among other things, helped transform the three-factor NEO inventory into the five-factor NEO-PI-R (McCrae & Costa, 1985), personality psychology accorded insufficient attention to many aspects of these broad factors.

The lay conceptions uncovered in natural language analyses suggest necessary but not sufficient features for a scientific classification, that is, features too important to leave out (Goldberg & Saucier, 1995; Saucier & Goldberg, 1996b). The lexical hypothesis assumes that (a) the most distinctive, significant, and widespread phenotypic attributes tend to become encoded as single words in the conceptual reservoir of language and (b) the degree of representation of an attribute in language tends to correspond with the relative importance of the attribute (Saucier & Goldberg, 1996b; cf. Osgood, May, & Miron, 1975, p. 45).

Both the utility and the limits of lexical studies have been acknowledged by the authors of the NEO facets:

> The body of trait terms in the English language provided a useful guide to broad factors, but many important distinctions that trait psychologists make have not been encoded in lay adjectives (McCrae, 1990), so analyses of adjectives are unlikely to be sufficient for identification of facets. (Costa et al., 1991, p. 887)

It is certainly true that expert distinctions can extend beyond lay conceptions and that lexical studies are not a sufficient criterion for identifying subcomponents. However, important distinctions, even expert ones, eventually tend to become encoded as English-language adjectives (Goldberg & Saucier, 1995; Saucier, 1992b), and lexical studies can lead to a minimum set of subcomponents necessary for any comprehensive hierarchical model of personality attributes. Therefore, lexical studies can potentially provide a useful check on the comprehensiveness of any theorized classification such as that of the NEO-PI-R. In this respect, lexical studies can yield indispensable content validity criteria. A researcher might fruitfully ask, Based on lexical studies, what is being unduly left out of this purportedly comprehensive personality inventory? In transforming their inventory from a three-factor to a five-factor measure, McCrae and Costa (1985) provided an appropriate response to exactly this question.

But perhaps this process of content-validity enhancement can go farther than the broad factors. There is no basis for drawing an arbitrary line between factors and their subcomponents, arguing that lexical data are relevant to one but irrelevant to the other. Admittedly, lexical studies do not have an unlimited range of usefulness. As mentioned, the function of lexical studies is to identify necessary rather than sufficient features, and natural-language person descriptors refer mainly to broad-level (e.g., is sociable) rather than highly specific (e.g., loves costume parties) behavior patterns. Therefore, lexical data should prove less and less relevant as one progresses to more and more specific subcomponents. Indeed, the lexical approach is ultimately self-limiting in this regard: The more specific the subcomponent, the less likely its replication across data sets in diverse languages. Accordingly, the proper role of lexical studies is to identify necessary but still fairly broad-level features of a comprehensive model of personality; it cannot identify all features of such a model.

AB5C Segments

The most prominent lexical studies of Big Five subcomponents have used a framework called the abridged Big Five dimensional circumplex (AB5C) model (Hofstee, De Raad, & Goldberg, 1992). The AB5C model conceives of subcomponents in a circumplexical rather than hierarchical manner, that is, in a horizontal rather than a vertical manner (Goldberg, 1993a). Because those terms denoting personality attributes tend to load on more than one factor, many of them can be seen as "blends" of factors (Saucier, 1992a). In the AB5C model, each term is plotted on the circumplex formed by the pair of Big Five factors on which it has its highest loadings and is characterized in terms of this pair. For example, *dominant* has its highest loadings on Factors I+ (Extraversion) and II− (Disagreeableness) and thus falls on the I–II circumplex. Because Big Five factors can be paired in 10 combinations, 10 such circumplexes are posited.

In the AB5C model, each circumplex plane is partitioned into segments of 30 degrees each, with a total of 90 segments included in the entire model. Therefore, each Big Five factor is composed of nine subcomponents (or segments) at each pole, including the factor-pure facet (e.g., for the I+ factor pole: I+I+, I+II+, I+III+, I+IV+, I+V+, I+II−, I+III−, I+IV−, and I+V−). For example, *dominant,* because its highest loading is on I+, falls into the I+II− segment, along with quasi-synonymous terms such as *opinionated, domineering, boastful,* and *forceful*; this segment is opposite to the I−II+ segment, which includes the related terms *modest, submissive, timid, and unaggressive* (see Hofstee et al., 1992, Figure 1).

Although this geometric scheme for conceptualizing subcomponents is extremely elegant, it has some limitations. It would be quite astonishing if there were exactly the same number of meaningful subcomponents, and exactly nine of them, for each of the broad personality factors; again, nature is not typically so neat. Indeed, one might argue that if these broad factors are truly orthogonal, all AB5C segments should be filled with descriptors. However, it may be that lay perception is imperfectly differentiated from a general evaluation dimension (e.g., Saucier, 1994). A key function of the expert may be to fill in unfilled segments, which will often be evaluatively inconsistent blends such as I+V− and III+IV−. Analyses have indicated that a number of the AB5C segments are virtually devoid of common person-descriptive terms. With regard to the set of 540 adjectives analyzed by Hofstee et al. (1992), 6 of the 90 segments contained none of these terms, and 27% (24 segments) contained two terms or less; at the other extreme, 18% (16 segments) contained from 10 to 24 terms.[2] Similarly, when Saucier (1992a) analyzed the factor-loading patterns of personality adjectives, he found that blends occurred most often in the planes defined by Factors I, II, and IV and decidedly less often in those planes involving Factor V. Moreover, evalu-

[2] In unpublished analyses, Goldberg generated similar assignments to AB5C categories for a larger set of 1,710 adjectives. In this data set, segments contained from 3 to 50 (rather than 0 to 24) terms, indicating again that some segments are more important than others in English-language analyses. Nor is English unique: When Hofstee, Kiers, De Raad, Goldberg, and Ostendorf (1997) attempted to find representative terms for all 90 segments in Dutch and in German, they reported no Dutch term for 14 segments and no German term for 18 segments.

atively consistent blends (e.g., I+V+) are more common than evaluatively inconsistent ones (e.g., I+V−). Clearly, natural-language terms are unevenly distributed across the segments, and this sword cuts both ways. On one side, it suggests that AB5C does not contain bona fide circumplexes. On the other side, the examination of this issue is made possible only because the AB5C approach allows identification of the more and the less heavily represented areas in descriptive space.

A second limitation of the AB5C approach is that segments can be populated by terms that do not necessarily form a coherent set. For example, the segments V+II+ and V−II−, respectively, contain the sets "deep, idealistic, diplomatic" and "terse, shallow." Terms that fall together in one segment are not necessarily highly intercorrelated; they simply share a common pattern of correlations (sometimes low) with a pair of broad factors, and this pattern may itself be difficult to replicate. One possible reason for the incoherence of the terms found in some AB5C segments is sampling error. Factor loadings and the covariances from which they derive are parameter estimates, and the accuracy of such estimates tends to increase with sample size: Four- and possibly five-digit samples may be required to make the segment assignments (and thus the groupings) stable from one sample to another.

Goldberg's 100 Clusters

Goldberg (1990) offered an alternative set of 100 specific subcomponent "clusters" that, when appropriately grouped, provide a measure of the Big Five factors. The clusters consist of adjectives joined because of their empirical relations to one another (alphas range from .41 to .86) rather than to a few broad factors; thus, the groupings are more consistently coherent than those produced by the AB5C model. The same number of subcomponents is not associated with each factor; Factors I to V include 25, 32, 23, 9, and 11 clusters, respectively. But there are an arbitrary number of subcomponents (i.e., 100 clusters), and the clusters were derived by expert procedures difficult to replicate in an exact manner (Block, 1995). Indeed, as a hierarchical model, the 100 clusters lack many of the advantages of the AB5C model: Not only are the derivation procedures less repeatable, but secondary factor loadings and circumplexical relations between variables are ignored. Moreover, scoring the 100 clusters requires the administration of a lengthy set of 335 adjectives (Goldberg, 1990). Because of the limitations of this model and other previous models, it seemed useful to develop a robust, empirically based, alternative set of subcomponents for the Big Five.

Hypotheses

We began with several specific hypotheses. First, we predicted that some subcomponents, at least two for each factor, could be replicated by lexical methods. Second, we predicted that the number of replicable subcomponents would differ according to the factor. Third, we predicted that the replicable subcomponents would fall in an orderly manner into those AB5C segments (e.g., II+IV+, I+II+, II−III−, and I+IV+) that tend to be most populated with descriptive terms.

The studies described here used the Big Five as a reasonable starting point for identifying subcomponents, which was sensible given the wide array of evidence supporting that model. Nonetheless, recent investigations indicate that there may be more robust solutions of fewer than five factors (Di Blas & Forzi, 1998; Digman, 1997; Saucier, 1997, 1998). Because these two- and three-factor solutions cover essentially the same descriptive terrain as the Big Five—although with factors of somewhat broader bandwidth—we would expect to be able to fit our findings easily into models of fewer factors than five.

Study 1

Method

Guidelines for variable selection. We sought variable selections involving as large a number of adjectives as practical while omitting relatively unfamiliar terms and having a reasonable comparability between the American and German adjective pools with respect to both the kinds of content excluded and the number of terms. We did not, however, seek a set limited to equivalent terms in the two languages (as we might have were we studying cultural differences). Rather, we retained whatever differential semantic emphases either language might contain and allowed their expression in language-specific "emic" factor structures; we then determined whether there were equivalences at the level of clusters of terms rather than at the more specific level of single terms in isolation. Although this approach limited our precision in identifying cultural differences, it maximized our power to identify cross-language universals.

American data set. We used the same data sets used by Goldberg (1990, Study 2, Samples A and C) to derive Big Five clusters. Three hundred twenty college students described themselves using a 587-adjective inventory with a 7-point response scale. Of these students, 316 used the same inventory to describe someone of their sex and approximate age they knew well and liked. Because we wanted our results to be based on the most precise parameter estimates possible and to generalize to both self-ratings and peer ratings, we pooled the self-ratings and peer ratings to form a large single sample (N = 636).

For the present analyses, the original set of 587 adjectives was reduced by eliminating (a) 47 "amplification" terms (e.g., *overcasual* and *oversentimental*), (b) 22 peripheral terms added to the 587 set at a late stage, and (c) those 18 terms among the remaining 518 with the lowest means in familiarity ratings (Saucier, 1997). The reduction to 500 terms minimized problem terms while keeping content coverage suitably wide. To provide coordinates for the Big Five personality factors in this data set, we used factor scores from Goldberg's (1992) 100 unipolar Big Five markers, all of which were included in the stimulus set of 587 terms.

German data set. We used the same data set used by Ostendorf (1990a). Newspaper announcements were used to recruit 414 volunteers from a medium-sized city and its environs in northwestern Germany. Participants were 170 men and 239 women who identified their sex; their average age was 33 years (range: 15 to 81 years). Each participant was sent an inventory of 830 adjectives for self-description. Each participant was also sent a second inventory containing the same 830 adjectives and instructed to pass this inventory along to one or more acquaintances or friends who would describe the target participant and mail the form back to the researcher. These peer-rating forms were returned for 394 of the targets; the reported average length of acquaintance was almost 11 years.

In the present study, we included only those participants with a very low frequency of missing responses. For the 393 self-ratings and 382 acquaintance ratings that were retained, the few remaining instances of missing data were converted to middle (neutral) responses. In instances in which more than one acquaintance returned rating forms for a target participant (this occurred for about 43% of the targets), the ratings were averaged. Ostendorf (1990a) found little difference in the factor structure produced by self-ratings and acquaintance ratings; to increase precision and generalizability, we pooled self-ratings and acquaintance ratings in all of our analyses to form a large single sample (N = 775).

Ostendorf's (1990a) inventory of 830 adjectives included 430 adjectives identified as prototypical disposition terms (either prototypical temperament–character traits or prototypical abilities–talents), based on classifications by 10 judges (described by Angleitner, Ostendorf, & John, 1990), as well as additional adjectives used in previous Big Five and interpersonal circumplex measures or otherwise deemed relevant on the basis of a review of the literature. The 430-item selection of disposition terms was somewhat narrower than that in Goldberg's set of 540 adjectives. In particular, it featured fewer terms referring to affective traits; judges tended to be divided on whether to classify such terms as dispositions or as temporary conditions. Accordingly, we added to the set of 430 an additional 132 adjectives so as to include all terms among the 830 that at least half (5 of 10) of the judges had classified as disposition terms (both traits and abilities), provided that even more of the judges had not classified the term instead as temporary conditions, social and reputational aspects, or overt characteristics and appearance. This led to a larger set of 562 terms better resembling the American variable selection. To arrive at a set of a size comparable to the American selection, we then eliminated those 62 terms that had elicited no response among at least 15 of the original 808 participants and targets, suggesting relative unfamiliarity.[3] Of the 430 terms used earlier by Ostendorf (1990a), 380 were included in our final set of 500 terms.

To provide coordinates for the Big Five personality factors in this data set, we used those 100 adjectives (20 for each factor) identified as having the highest loadings on the Big Five by Ostendorf (1990a, Tables 52 through 56) to generate Big Five factor scores.

Analyses. The same analytic procedures were followed in each of the two data sets. First, factor scores (principal components) were generated from the 100 Big Five marker items. Next, these factor scores were correlated with the 500 adjectives. As an initial inclusion rule, we selected all adjectives with a correlation of at least .25 with one of the five factors; each of these terms was assigned to one of the five domains on the basis of its highest correlation.

There could be two sources of error in assignments to factor domains based on this first inclusion rule: (a) errors in the pattern of correlations with the factors, holding the factor axes constant, and (b) variations in the factor axes from one data set to another. Pertinent to the first type of error, an examination of fluctuations in AB5C angular locations between two large American data sets showed an average fluctuation of 13.5 degrees across 257 common variables. With regard to the second type of error, analyses conducted by Hofstee, Kiers, De Raad, Goldberg, and Ostendorf (1997) indicated impressive congruence between the Big Five factors derived from Goldberg's 100 markers and from Ostendorf's 430 adjectives for Factors I, II, III, and V. The German Factor IV had a lower correlation with its American counterpart, largely because it was more specific and less diffuse than the American one.[4] Thus, errors of the second type, although inevitable, seemed generally small.

Nonetheless, were nothing done to correct for the errors just described, correspondence across data sets could be reduced simply because of fluctuations in the variable loadings and the locations of the factor axes. That is, corresponding variables could conceivably be assigned to noncorresponding domains in the two data sets. To provide some correction for these errors, we assigned adjectives to a factor domain so long as they were within 55, rather than just 45, degrees of the factor axis. This criterion translates into the following final inclusion rule: Any adjective with a correlation of at least .25 with one of the five factors was retained within that domain, providing that correlation was either (a) the highest with any of the five factors or (b) of a magnitude at least 70% as high as its highest correlation with any of the other four factors.[5]

For each factor domain within each data set, correlations between the adjectives were analyzed via principal components with oblique rotations. We inspected rotations of 2, 3, 4, 5, and 6 factors for each factor domain using the promax method (Hendrickson & White, 1964), an oblique-rotation algorithm based on the well-known varimax procedure, with the kappa parameter (exponent) set at the common default of 4.[6] Oblique rotations are more appropriate in exploratory analyses that aim at specific factors in a set of variables known to be highly related, because one would therefore expect correlated factors (Goldberg & Digman, 1994).

Results

Each of the two sets of 100-adjective Big Five markers produced the Big Five with high clarity, with the highest-loading terms resembling those in published analyses (Goldberg, 1992; Ostendorf, 1990a). Use of the inclusion rules described earlier led to 121, 121, 77, 70, and 35 adjectives for Factor Domains I through V in the American data set and to 120, 162, 126, 62, and 115 adjectives for the corresponding factor domains in the German data set. The most striking difference between the American and German selections was the much larger number of adjectives in the Factor V domain in the German data set, probably resulting from the inclusion of more terms denoting abilities and talents in that data set.[7]

There were some clear differences between the German and American facet structures. The American structures contained

[3] A number of these terms were cognate terms imported from other languages into German but not in common use (e.g., *Altruistisch, Cholerisch, Egozentrisch, Exzentrisch, Generoes,* and *Inkompetent*).

[4] For Factors I, II, III, and V, Tucker's phi coefficients for loadings on 126 corresponding (translated) terms were, respectively, .80, .82, .75, and .79. For Factor IV, the coefficient was .58; the American Factor IV also had substantial phi coefficients with other German factors (.44 with German Factor III, .43 with German Factor V, and .38 with German Factor II), giving it a much higher multiple correlation with factors in the other language than found for the German Factor IV and suggesting that the American data involve a broader Factor IV. These figures also are consistent with the conclusion that the American Factor IV is more evaluative than the German one, because these are correlations with those German factors (II, III, and V) having stronger evaluative valences. Relative to the American Factor IV, the German Factor IV appeared to be rotated 15 to 20 degrees away from Factors II and V. Rotation in the other planes appeared to be much smaller.

[5] The value .70 corresponds to 1/tan for 55 degrees, whereas the corresponding 1/tan value for 45 degrees is 1.00. At 45 degrees, the ratio of the pair of highest loadings is 1:1; at 55 degrees, however, the ratio is 7:10.

[6] In the promax procedure, varimax factor loadings provide an initial factor matrix that is then rotated by the Procrustes procedure to a solution of even simpler structure with the orthogonality constraint relaxed. The effect of the promax rotation is to allow factor axes to rotate from their orthogonal varimax positions back toward the first unrotated principal component (Cureton & Mulaik, 1975), which in each of these 10 data subsets would be the relevant Big Five factor axis. We also generated varimax factors to compare with the promax factors. For the most closely corresponding factors, factor scores from varimax and promax procedures correlated .83 and above in every case; 182 of the 200 correlations were .90 or higher, and only one was under .86. Interestingly, the varimax and promax positions tended to be slightly closer in the American data, because the mean interadjective correlations within factor domains were lower in the American data.

[7] Examples of Factor V German terms (as translated in Ostendorf, 1990b) having no near equivalent in the set of English adjectives are reflective, original, capable, incompetent, unartistic, uninformed, and stupid. Further examples include Factor V German terms having cognate forms in English that were not in the English adjective selection: musical, unmusical, talented, untalented, talentless, and cultivated.

facets focused on courtesy–politeness (Factor II) and contemplativeness (Factor V) that had no apparent counterpart in the German data. The German structures contained facets focused on wittiness (Factor I), honesty (Factor II), and need for recognition (Factor IV) that had no apparent counterpart in the American data. Generally, the German Factor I (Extraversion) contained a stronger impulsivity subcomponent than did the American Factor I.

Although there were some differences between the German and American facet structures, there were also many similarities. There is evidence of robust Big Five facets when one compares the corresponding American and German structures. Such a comparison suggested many content themes within the Big Five facets in both languages. For example, in both languages, the facets of Factor I differentiated along the themes of sociability, confidence–boldness, exhibition–disinhibition (i.e., not being quiet or reserved), and daring–adventurousness.

These correspondences suggest that indeed the lexical approach can help in identifying necessary features for a model of personality description at a finer grained level than the broad Big Five level. But caution is in order. Such informal identifications of common content themes in adjectives from two different languages are prone to subjective biases. Study 2 established a more objective basis for comparing subcomponents across the two languages.

Study 2: A Subcomponent-Comparison Study

Stronger evidence for robust lexical subcomponents of the Big Five would be provided by data from German–English bilinguals. The bilinguals should be drawn from native speakers of both languages and from inhabitants of both German-speaking and English-speaking nations to control for the effects of first language and current social setting. Ideally, one would ask such a group of bilinguals to judge the applicability of the terms to a set of specific behaviors. Also valuable would be data provided by a large sample of German–English bilinguals in Germany and the United States describing real people in both languages. Unfortunately, such samples are difficult to obtain. Also pertinent, but far more practical and economical, are the judgments by a group of bilinguals as to the correspondence of the respective German and American subcomponent structures.

How might this correspondence be judged? One way to compare factors across a language boundary is to translate the items that load most highly on the factors. Translations facilitate comparative judgments by monolingual observers but raise thorny issues. Taking a pessimistic view, Hofstee (1990) argued that "the majority of trait terms are not adequately translatable," primarily because of discrepancies in the "social desirability of terms with corresponding denotation" (p. 81). For example, Dutch and English have Latin-root cognates for *aggressive* and *critical*; in Dutch, the former is undesirable and the latter is desirable, whereas American English shows the opposite pattern. On the other hand, some studies (Kosmitzki & John, 1993; Peabody, 1985) suggest a more optimistic view of the translatability of personality terms.

In this study, we averted the difficult issues associated with translation by recruiting a sample of German–English bilinguals who were able to compare the subcomponents from American and German data without translation. Because these judges compared aggregates of terms, many of the problems associated with establishing the exact commensurability of a pair of single terms (given the potential ambiguity of any single term taken alone; Goldberg & Kilkowski, 1985) were attenuated.

Method

Raters. We recruited 22 raters fluent in both German and English, 11 in the United States and 11 in Germany; most were university students. There were 9 female and 13 male judges, and the mean age of the judges was 28 years. Fourteen of the judges were natives of Germany. The German natives had spent an average of 9 years learning English in school, and the nonnatives reported an average of 6 years learning German in school. All German-native judges had spent at least 6 months (and as many as 26 years) living in an English-speaking country, and all of the other judges had spent at least 2 years (and as many as 24 years) living in a German-speaking country. Fluency of the judges was assessed solely on the basis of their schooling and experience abroad.[8]

Materials. Raters were presented with forms containing instructions in both English and German.[9] For each promax factor, we provided the six English (or six German) adjectives having the highest loadings. Judges were instructed to indicate which English cluster provided the best match to each German cluster and to rate the relative closeness (1 = *weak,* 2 = *moderate,* 3 = *very close*) of the match they had made.

Analyses. At each level of the promax factors (2, 3, 4, 5, and 6), a matrix of German × English adjective-cluster judgments was computed. Each entry in the matrix consisted of the averaged scores of raters as to the closeness of the match; when two adjective clusters were not chosen as a match, the rater's judgment was always scored as zero. One reasonable standard would define replication as an average score of 2.00 or more (i.e., 67% of the highest possible score), indicating an average judgment that the clusters were at least moderately well matched. This criterion could be reached by having all judges match a German with an English subcomponent with a moderate (2.00) rating or by having two thirds of judges make the match at the highest possible (3.00) rating. This 67% cutoff is, however, arbitrary. We therefore examined whether the data suggested a more natural cutoff.

Results

How many subcomponents are replicable? Although between-judges agreement varied widely from one subcomponent to another, it was generally at a high level; on average, pairs of judges agreed on 63% of their matches. This high percentage suggests substantial reliability in the aggregated judgments.[10]

Table 1 presents the degree of matching, by the percentage index, for each of the subcomponents. Overall, the number of subcomponents judged to be replicated would necessarily depend on the point at which one sets the replication cutoff. As Figure 1

[8] It is possible that a more objective assessment of bilingual fluency would have improved our sample of judges and, perhaps, reduced the level of error and chance variation in the results, because a group of nonfluent judges should produce little more than chance levels of agreement. However, such a reduction in noise would be unlikely to change the nature of the signal.

[9] Copies of these instructions are available from Gerard Saucier.

[10] Within subgroups of the judges defined by male–male pairings, female–female pairings, male–female pairings, pairings of native Germans with native Germans, pairings of nonnative Germans with nonnative Germans, or pairings of native Germans with nonnative Germans, mean percentage agreement rates varied from 58% to 64% and gave no clear evidence that these groups produced divergent results.

Table 1
Percentage Matching Index for Factors in Big Five Domain Solutions

Factor	Percentage index						No. above 55% criterion
	1	2	3	4	5	6	
I (Extraversion)	**89**	**80**					2
	89	**82**	**76**				3
	92	**85**	**83**	**76**			4
	79	52	45	29	12		1
	59	48	41	38	32	29	1
II (Agreeableness)	**77**	42					1
	88	**67**	35				2
	86	**64**	32	14			2
	88	50	45	29	20		1
	91	**83**	**61**	**55**	27	15	4
III (Conscientiousness)	**56**	33					1
	59	38	24				1
	73	52	48	30			1
	86	**83**	**73**	**68**	44		4
	97	**68**	**65**	36	27	21	3
IV (Emotional Stability)	**89**	**79**					2
	91	**89**	**80**				3
	70	**61**	**57**	44			3
	61	47	44	36	20		1
	53	52	44	29	27	17	0
V (Intellect–Imagination)	**98**	**94**					2
	95	**83**	33				2
	97	**58**	45	44			2
	79	**58**	41	36	23		2
	64	**56**	**56**	44	26	17	3

Note. For each Big Five domain, solutions are listed in order, from two-factor to six-factor solutions, with percentage matching indexes for that solution listed in descending order of magnitude. Percentage indexes of 55 or more are in boldface.

indicates, when we began with a high (95%) criterion and progressively relaxed the criterion by 5% increments down to the 55% level, each 5% increment led to at least one additional (distinct from those already included) replicated subcomponent. But from the 55% level, it was necessary to lower the criterion by a full 15% (to 40%) to obtain further replicated subcomponents. Therefore, the 55% criterion was judged to be the least arbitrary cutoff. This 55% cutoff led to a total of 18 replicated subcomponents.

For Factor I, the two-factor, three-factor, and four-factor structures were all replicated. Thereafter, in the five-factor and six-factor structures, replication dropped off dramatically. Table 1 provides a clear indication that four Extraversion subcomponents were replicated across two languages. These replicated subcomponents can be labeled as sociability, unrestraint, assertiveness, and activity–adventurousness.

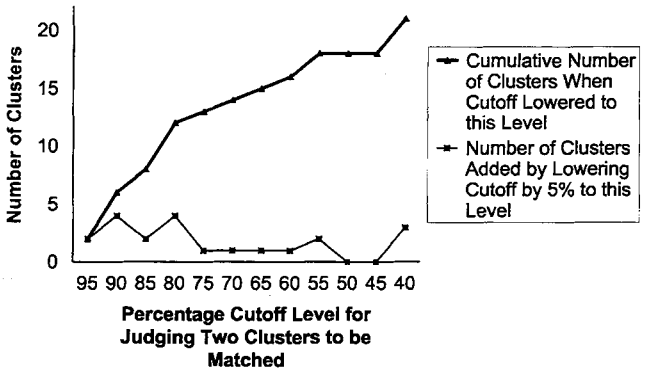

Figure 1. Number of clusters added as a function of cutoff level.

Table 1 also provides clear evidence for four replicated Conscientiousness subcomponents. In the five-factor structure, four of the five factors replicated. The replicating subcomponents could be labeled as orderliness, decisiveness–consistency, reliability (or trustworthiness), and industriousness. The nonreplicating fifth subcomponents (in the five-factor solutions) were self-control in the German data and practicality–carefulness in the American data.

Table 1 provides evidence for three replicating Emotional Stability subcomponents. The entire two- and three-factor structures were replicated. The replicating subcomponents could be labeled more easily in negative terms, as irritability, insecurity, and emotionality.

For Factor V, Table 1 indicates a replicating two-factor structure. The replicating subcomponents could be labeled intellect and imagination–creativity. In the six-factor solutions, three factors replicated, two of which were also intellect and imagination. The third subcomponent had a perceptiveness–farsightedness theme. Indeed, a perceptiveness subcomponent was found in both languages in solutions of four or more factors.

The situation with Factor II (Agreeableness) was more complex. In the two-factor solutions, a warmth–affection subcomponent replicated. In the three-factor solutions, both warmth–affection and civility (considerate–cooperative vs. rude–antagonistic) subcom-

ponents replicated. In both languages, beginning with the four-factor solutions, the civility subcomponent was subdivided into narrower bandwidth factors. In the four-factor solutions, one aspect of civility (gentleness, or tolerant vs. harsh) was replicated. In the five-factor solutions, warmth–affection was the only replicating subcomponent. Finally, in the six-factor solutions, there were four replicating subcomponents: warmth–affection, gentleness, generosity (or unselfishness), and humility–modesty (vs. conceit). The last three of these components were related to the broader civility subcomponent in the three-factor solutions.

For Agreeableness, the American data produced cheerfulness and politeness subcomponents that had no close match in the German data, whereas the German data yielded slyness and honesty subcomponents with no close match in the American data. As noted earlier, because the design of our studies was oriented toward identifying cross-language commonalities rather than differences, we cannot be sure of the origin of these discrepancies. Some or all of these differences may have been a simple function of differing variable selection procedures. For example, several German terms referring to impoliteness were classified as temporary states and thus omitted from the German variable selection.

However, it is noteworthy that Agreeableness and, to a lesser degree, Conscientiousness appeared to have the greatest cross-language variation in their subcomponents. Within each language, both of these factors seem to be rich in meaningful subcomponents, but the division into subcomponents found in one language seemed to have relatively less commonality with that in the other language relative to the situation found for Emotional Stability, Intellect, and, especially, Extraversion. The attributes implicated by these factors—desirable ones such as warmth, generosity, cordiality, neatness, decisiveness, reliability, and industriousness and undesirable ones such as coldness, stinginess, antagonism, sloppiness, and laziness—may form a distinct class of traits. They concern vital aspects of "love and work" and seem to be the factors most highly associated with conceptions of ethics, virtue, and character. In these domains, different languages (and the cultures these languages reflect) may be especially prone to diverge, either in value emphases or in hierarchies of conceptualization. It would seem, then, that conceptualizations within the domains of Agreeableness and Conscientiousness may be especially fertile for studies of cultural differences.

Summing over the five domains (i.e., the 4 replicating subcomponents each from Factors I, II, and III and the 3 replicating subcomponents from Factors IV and V), we found 18 replicating subcomponents of Big Five factors. These subcomponents replicated across languages, but how well were they replicated across self-ratings and peer ratings within each language? Recall that, to enhance generalizability, we combined self and peer data in Study 1 to derive subcomponents. We took a retrospective look at the levels of within-language self–peer replication for these 18 subcomponents.

In the American data, 11 of the 18 subcomponents were virtually identical in the self and peer data, with both self and peer factors correlating .96 or higher with the corresponding combined data set factors. For 5 further subcomponents (activity–adventurousness, gentleness, humility—modesty, decisiveness–consistency, and perceptiveness), the correspondence was almost as high; both self and peer factors correlated .86 or higher with the corresponding combined data set factors. The 2 remaining subcomponents were less stable across American self and peer data. Generosity factors correlated .94 (self) and .68 (peer) with the combined data factors, whereas industriousness factors did not emerge clearly in either American self or peer five-factor solutions in the Conscientiousness domain. An industriousness factor emerged only when the two data sets were combined: Industriousness items distributed themselves across other self factors or peer factors.

In the German data, there were potentially greater dependencies introduced by having most participants included twice, once for self-reports and once as the target of a peer report. However, these dependencies seemed to be of no effect: In general, the factors generated independently in self-rating and peer rating data were virtually identical. All but 3 of these 36 factors correlated .96 or higher with the factor found in the combined data. The exceptions were the peer activity–adventurousness subcomponent (four-factor solution), which correlated .91 with the combined data set factor, and the perceptiveness subcomponent (six-factor solution), in which both the self (.66) and peer (.85) versions had a fairly moderate correlation with those from the combined data. However, in the four-factor solutions, the perceptiveness subcomponents from the self and peer data correlated .97 and higher with those in the combined data four-factor solution.

Overall, factor replication between self-ratings and peer ratings was very impressive, but a few subcomponents appeared more replicable across languages than across self-ratings and peer ratings within one language. How could this be? Factor stability depends to a degree on sample size (Guadagnoli & Velicer, 1988), so it should not be surprising that some factors that are not cleanly replicated within language in samples of about 300 prove to be stable across languages when the comparison is between samples of at least twice that size. However, this finding does imply a somewhat lower level of robustness for industriousness in particular, as well as for perceptiveness, generosity, and activity–adventurousness, relative to the other 14 subcomponents.

The content of the expanded set of subcomponents is represented in Table 2, which provides a set of adjectives to measure each of the 18 subcomponents, along with estimates of their reliabilities. Investigators wishing to test, replicate, or extend our findings might pool the 142 adjectives in Table 2 into an inventory, using instructions and a rating scale similar to that for Goldberg's (1992) 100 unipolar markers.

Table 3 shows the factor structure of the 18 scales. The correlation matrix of the 18 scales yielded the Big Five as latent variables. The first five eigenvalues were 5.1, 2.7, 2.0, 1.8, and 1.3. The next five were considerably lower: 0.7, 0.6, 0.5, 0.5, and 0.4. The first five accounted for 72% of the variance. These five varimax-rotated factors correlated .94, .86, .90, .90, and .89 with corresponding factors (I, II, III, IV, and V, respectively) derived from a set of Big Five marker adjectives (Goldberg, 1992).

A hierarchical structure such as that provided by these scales provides a suitable model for use in confirmatory analyses. However, model specification should take account of the higher of the secondary loadings in Table 3, probably all of those roughly .20 or higher in magnitude.

Table 3 also provides the structure of two- and three-factor rotations, given recent studies (Di Blas & Forzi, 1998; Digman, 1997; Saucier, 1997, 1998) indicating the robustness of such solutions. The two-factor solution followed that of Digman (1997):

Table 2
Lexical Big Five Subcomponents: English-Adjective Items and Reliability

Label and constituent adjectives	Alpha		MIC	
	Self	Peer	Self	Peer
Factor I				
Sociability (sociable, cheerful, merry, effervescent vs. unsociable, withdrawn, uncommunicative, seclusive)	**.81**	**.86**	.35	.44
Unrestraint (talkative, verbal, aggressive, domineering vs. quiet, untalkative, shy, reserved)	**.85**	**.85**	.42	.41
Assertiveness (assertive, straightforward, direct, bold vs. weak, cowardly, submissive, helpless)	**.76**	**.81**	.29	.35
Activity–adventurousness (daring, adventurous, active, competitive, rambunctious vs. unadventurous, uncompetitive, unenergetic)	**.79**	**.79**	.32	.33
Factor II				
Warmth–affection (affectionate, sentimental, sensitive, warm, compassionate vs. cold, unsympathetic, insensitive)	**.84**	**.87**	.40	.46
Gentleness (agreeable, cordial, amiable vs. harsh, antagonistic, hard, rough, combative)	**.72**	**.80**	.24	.33
Generosity (charitable, helpful, generous, unselfish vs. uncharitable, greedy, stingy, selfish)	**.70**	**.66**	.26	.23
Modesty–humility (modest, humble vs. egotistical, boastful, conceited, snobbish, vain, egocentric)	**.78**	**.84**	.30	.40
Factor III				
Orderliness (organized, orderly, neat, meticulous vs. disorganized, disorderly, sloppy, unsystematic)	**.90**	**.90**	.51	.52
Decisiveness–consistency (decisive, firm, consistent, steady vs. indecisive, inconsistent, scatterbrained, illogical)	**.73**	**.78**	.25	.30
Reliability (reliable, dependable, responsible, prompt, punctual, respectful vs. undependable, unreliable)	**.84**	**.89**	.43	.51
Industriousness (ambitious, industrious, purposeful, conscientious vs. aimless, negligent, lazy, unconscientious)	**.73**	**.80**	.26	.34
Factor IV				
(Low) irritability (undemanding, uncritical, tranquil vs. irritable, temperamental, moody, impatient, defensive)	**.72**	**.77**	.24	.29
(Low) insecurity (relaxed, unenvious vs. insecure, unstable, nervous, fretful, envious, jealous)	**.78**	**.75**	.31	.27
(Low) emotionality (unemotional, unexcitable vs. high strung, emotional, excitable, anxious, fidgety, suggestible)	**.70**	**.67**	.23	.21
Factor V				
Intellect (intelligent, intellectual, philosophical, analytical, knowledgeable, complex vs. unintellectual, unreflective)	**.79**	**.79**	.33	.33
Imagination–creativity (creative, inventive, imaginative, artistic, clever, innovative vs. uncreative, unimaginative)	**.87**	**.86**	.45	.44
Perceptiveness (perceptive, insightful, foresighted vs. imperceptive, unobservant, shortsighted)	**.64**	**.73**	.23	.31

Note. n = 320 self-ratings; n = 316 peer-ratings. Alpha coefficients are in boldface. MIC = mean interitem correlation.

an alpha factor based on Agreeableness, Conscientiousness, and Emotional Stability and a beta factor based on Extraversion and aspects of Intellect–Imagination.

AB5C findings. Table 4 indicates that the 18 scales fell into 15 unique AB5C segments when the scales were correlated with standard Big Five markers (Goldberg, 1992). Three scales (gentleness, generosity, and modesty–humility) fell into the II+IV+ segment, and two scales (intellect and imagination–creativity) fell into the V+V+ segment.[11] These findings appear to provide support for the usefulness of AB5C segments in demarcating

[11] It is interesting that the heterogeneous AB5C facets are associated with Factor II and Factor V. In as-yet-unpublished analyses, Ostendorf has found that these two factors split when seven factors are rotated in the data set used in the present Study 1.

Table 3
Eighteen Big Five Subcomponents: Factor Structure

	Five factors					Three factors			Two factors	
Factor and subcomponent	I	II	III	IV	V	I	II	III	α	β
Factor I										
Sociability	**.72**	.44	.05	−.07	.11	**.65**	.33	−.02	.32	**.63**
Unrestraint	**.80**	−.11	−.11	−.09	.02	**.69**	−.19	−.15	−.19	**.73**
Assertiveness	**.70**	−.07	.23	.26	.27	**.76**	−.10	.35	.20	**.69**
Activity–adventurousness	**.69**	−.02	.16	−.05	.17	**.73**	−.02	.03	.07	**.73**
Factor II										
Warmth–affection	.20	**.69**	.16	−.35	.24	.31	**.78**	−.19	**.49**	.26
Gentleness	−.09	**.74**	.09	−.03	.16	−.03	**.74**	.06	**.59**	−.06
Generosity	.20	**.66**	.20	.15	.10	.20	**.60**	.26	**.66**	.15
Modesty–humility	−.18	**.63**	.26	.20	−.05	−.19	**.61**	.33	**.68**	−.25
Factor III										
Orderliness	−.05	.06	**.66**	−.01	.06	.10	.25	**.31**	**.40**	.05
Decisiveness	.16	.08	**.71**	.38	.14	.27	.21	**.68**	**.58**	.20
Reliability	.04	.28	**.72**	−.02	.11	.20	**.46**	.35	**.60**	.14
Industriousness	.28	.24	**.64**	.06	.34	**.48**	.41	.40	**.62**	.42
Factor IV										
Irritability (low)	−.02	.58	.03	**.59**	−.02	−.09	.37	**.49**	**.58**	−.14
Insecurity (low)	.32	.26	.15	**.70**	.10	.24	.11	**.64**	**.47**	.19
Emotionality (low)	−.27	−.17	.03	**.72**	−.05	−.33	−.27	**.63**	.10	**−.31**
Factor V										
Intellect	.01	.04	.13	−.01	**.85**	**.33**	.23	.12	.28	**.31**
Imagination	.23	.10	.04	.03	**.57**	**.42**	.18	.08	.23	**.40**
Perceptiveness	.18	.21	.32	−.01	**.58**	**.43**	.37	.20	**.46**	.39

Note. N = 636. Coefficients are factor loadings from principal-axis solutions with varimax rotation. The highest loading for each variable within a solution is shown in boldface.

specific trait domains but suggest three caveats about the AB5C approach. First, one AB5C segment may contain heterogeneous content. For example, II+IV+ was the AB5C location for 3 subcomponents with clearly differentiable themes. Second, the coefficients in Table 4 do not support the assumption, inherent in the AB5C model, that only the two highest loadings are of importance. Five of these scales had more than two loadings of .20 or higher, and 10 had more than two loadings of .15 in magnitude or higher. Third, in the AB5C model, 20 of the 45 sectors (44%) were evaluatively heterogeneous, that is, a blend of the desirable pole of one factor with the less desirable pole of a second factor; however, only 1 (6%) of the 18 subcomponents (nonemotionality; IV+II−) was evaluatively heterogeneous in this way. These Big Five subcomponents clearly underrepresent the evaluatively inconsistent (i.e., neutral) segments of the AB5C model.

We had predicted that the replicating subcomponents would fall into those AB5C segments with the largest concentrations of adjective descriptors. A perfect confirmation of this hypothesis would be produced if the 15 most populated segments, either in English or in German, were the same 15 segments into which the subcomponents fall in Table 4. However, only 11 of the subcomponents were located in the 15 most populated segments in English (as per Hofstee et al., 1992); the situation was roughly similar with the German equivalents (when promax factor scores were used to determine positions within AB5C). It was clear that several anomalies prevented a perfect confirmation of this hypothesis. The most striking concerned the AB5C segment II+III+, the second most populous segment in both English and German. No replicated subcomponent from the present study fell into this segment. In the American data, the four-, five-, and six-factor solutions all had a promax factor that fell into the II+III+ segment; adjectives such as polite, courteous, cooperative, and considerate all had high loadings (roughly .60) on these promax factors. In the German data, all solutions (from two factors up) likewise had a promax factor that fell into the II+III+ segment; adjectives that might be translated as honest and truthful (as opposed to lying and bragging) all had high loadings.[12] However, our judges seldom matched these English and German factors, and indeed politeness and honesty are not really synonymous. This provides further evidence that an AB5C segment can be heterogeneous with respect to content and that more than one distinct personality construct can be located in such a segment.[13]

Fortunately, there are alternatives to AB5C for pictorially representing the 18 subcomponents. One is the tree diagram produced by a cluster analysis; as one would predict, cluster analyses using between-groups, within-group, and nearest-neighbor methods, with correlations as the distance measure, showed an agglomeration of the 18 subcomponents (in English) into those Big Five factors from which they were derived. But tree diagrams do not show some of the more complex patterns of interrelation among

[12] The German terms were *Ehrlich, Wahrheitsliebend, Luegnerisch,* and *Grosstuerisch.*

[13] Note that if an "AB2C" model were constructed using the two-factor solution in Table 3, all six bipolar pairs of segments in the single AB2C circumplex would be filled by at least 1 of the 18 subcomponents; there would be no gaps. Lay perception seems to have differentiated enough to fill in even the evaluatively inconsistent segments in this model defined by two broader factors.

Table 4
Eighteen Big Five Subcomponents: Correlations With Big Five From Standard Markers

Factor and subcomponent	Big Five factor score I	II	III	IV	V	AB5C segment
Factor I						
Sociability	**.69**	**.47**	.07	.13	.10	I+II+
Exhibition–disinhibition	**.91**	.00	−.12	−.06	.00	I+I+
Assertiveness	**.66**	−.01	.20	**.25**	.24	I+IV+
Activity–adventurousness	**.64**	.14	.16	−.01	**.18**	I+V+
Factor II						
Warmth–affection	.13	**.82**	.10	−.15	.19	II+II+
Gentleness	−.15	**.69**	.08	**.19**	.15	II+IV+
Generosity	.12	**.66**	.15	**.34**	.08	II+IV+
Modesty–humility	−.24	**.50**	.24	**.32**	−.07	II+IV+
Factor III						
Orderliness	−.04	.01	**.87**	−.05	.01	III+III+
Decisiveness–consistency	.12	.08	**.68**	**.35**	.08	III+IV+
Reliability	−.02	**.36**	**.67**	.04	.05	III+II+
Industriousness	.23	.27	**.63**	.13	**.29**	III+V+
Factor IV						
Irritability (low)	−.11	**.36**	.05	**.74**	−.05	IV+II+
Insecurity (low)	**.25**	.06	.19	**.82**	.09	IV+I+
Emotionality (low)	−.25	**−.37**	.06	**.63**	−.08	IV+II−
Factor V						
Intellect	.04	.12	.20	.04	**.77**	V+V+
Imagination–creativity	.18	.06	.09	.12	**.80**	V+V+
Perceptiveness	.18	.26	**.29**	.05	**.50**	V+III+

Note. $N = 636$. Each subcomponent's highest loading is shown underlined and in boldface. Secondary loadings large enough to affect an AB5C segment assignment are shown in boldface. Assignments to AB5C segments are to 1 of 45 bipolar pairs of opposing segments, each segment spanning 30 degrees. AB5C = abridged Big Five dimensional circumplex.

variables. As an alternative, we offer Figure 2, which shows, with solid lines, all correlational linkages among the 18 subcomponents that exceeded .45 (in the American data). One subcomponent (Nonemotionality) had no correlation higher than .45; its two highest correlations are depicted with dashed lines. In Figure 2, one can discern the distinct groupings of variables that make up Extraversion (upper left), Agreeableness (lower left), Emotional Stability (lower middle), Conscientiousness (right), and Intellect–Imagination (upper right). One limitation is that some of the correlations under .45 violate the relative spatial locations shown; for example, although located far from them, generosity is correlated .37 with both reliability and industriousness. On the other hand, Figure 2 confers one advantage over other representations: It facilitates visualizing alternative ways of variously lumping and splitting the subcomponents. As one example, Digman's (1997) two higher order factors divide generally into the bottom (alpha) and top (beta) subcomponents in Figure 2.

General Discussion

We hypothesized that some subcomponents, at least two for each Big Five factor, could be replicated, and our results bore out the hypothesis. We also found, as hypothesized, that the number of replicable subcomponents would differ according to the factor; for each Big Five factor, by a strict criterion, we found either three or four subcomponents to be replicable across two languages.

The results have several consequences for future investigations of personality and individual differences. Most broadly, they indicate that the Big Five is only at the tip of the iceberg of personality structure and that lexical analyses can identify finer grained factors that have commonality across languages. Measures of the broad Big Five factors alone underspecify the domain of meaningful personality attributes. Finer grained attributes, such as the facet scales of the NEO-PI-R (Costa & McCrae, 1992) or the present subcomponents, provide higher fidelity.

Moreover, the present study made use of a potentially useful methodological innovation. Traditionally, factor analysts have sought to identify a dozen or more "primary" oblique factors, which might be placed in a hierarchical representation by use of higher order factor analysis of the factor intercorrelations. This "bottom-up" method has, however, not yet had fertile application in lexical studies, probably for two reasons: (a) Primary factors are not extremely stable even within one language, as a result of increasing capitalization on random error as one increases the number of factors, and (b) it would be more difficult to find the same set of primary factors in a cross-language comparison.[14] The present study proceeded in a reverse "top-down" direction, starting with the broad, independent factors and identifying their subcomponents. This top-down methodol-

[14] Cattell's work would be the possible exception. However, Cattell's classic studies (e.g., Cattell, 1943) do not come up to standards for contemporary lexical studies of personality, because of nonrepeatable procedures and many subjective rather than empirically based decisions (Block, 1995; John, 1990).

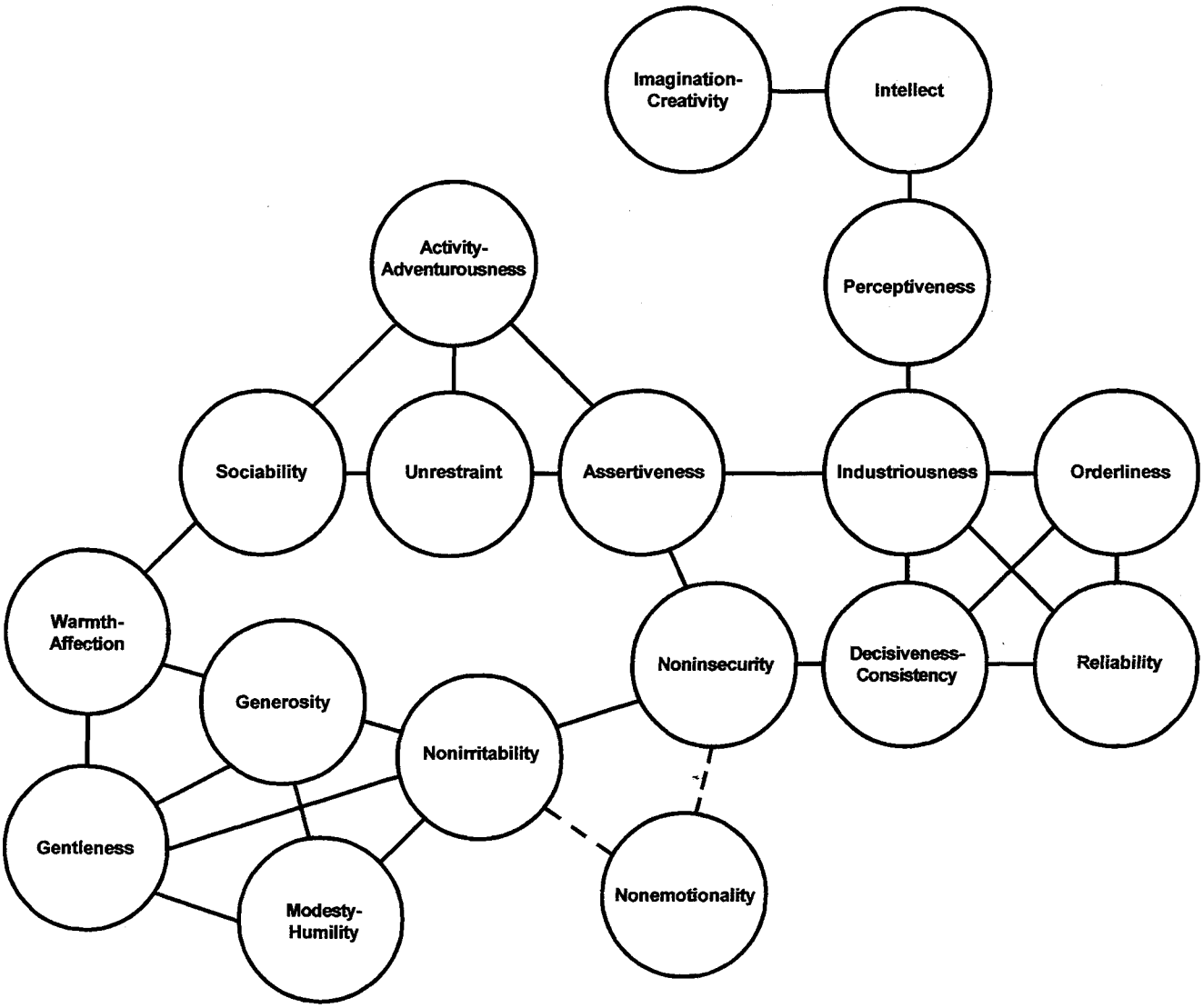

Figure 2. An alternative representation of the 18 subcomponents of the Big Five.

ogy has at least one pitfall: It does not allow outlier adjectives (e.g., Saucier & Goldberg, 1998) to contribute to subcomponents or to form their own subcomponents. Nonetheless, replicated factor structures seem more attainable when one demarcates a domain and seeks a few (five or less) factors rather than seeking many primary factors all at once. Therefore, the top-down approach will often produce a clearer pattern of results at the level of specific subcomponents, providing a useful complement to the more traditional bottom-up approach.

How are these findings useful? As argued at the outset, the lexical approach leads to valuable content-validity benchmarks. A personality factor found to be robust not only in a single language, but also across multiple languages, should not be ignored in a personality model or inventory that aspires to be comprehensive. Therefore, the subcomponents identified in the present study provide a way of assessing the comprehensiveness of broadband personality inventories, especially those—such as the NEO-PI-R (Costa & McCrae, 1992), the Sixteen Personality Factor Questionnaire (16PF; Cattell, Eber, & Tatsuoka, 1970), and the Personality Research Form (Jackson, 1984)—that combine broad and specific variables. Future studies might usefully determine whether content found in the present subcomponents is omitted in any of these inventories. The best index would be the multiple correlation of an inventory's entire set of scales with each of the present subcomponents. A high multiple *R* value would indicate that the inventory captures the content of the subcomponents, even if it does so in a complex manner. A low multiple *R* value would suggest omitted content. If content is, in fact, omitted the inventory might be revised to make it more comprehensive, just as comparison with the lexical Big Five led to a fruitful expansion of the early NEO inventory (McCrae & Costa, 1985).

With regard to the NEO-PI-R, our initial impression is that most of the content of the 18 subcomponents is found among the inventory's facet scales, but this impression needs empirical con-

firmation. The 16PF provides an even more intriguing comparison, because this inventory has approximately the same number of specific-level constructs as we found in our cross-language replication. Some 16PF constructs (e.g., self-reliance, vigilance, and rule consciousness) are difficult to reconcile with any of the present 18, and some of the latter (e.g., most of the Agreeableness and Conscientiousness subcomponents) are difficult to reconcile with any of the 16PF scales. It does not appear that the 16PF can be considered representative of replicated lexical hierarchical subcomponents, but we plan to assess this impression in future empirical work.

We would like to warn against two kinds of inappropriate inferences that might be drawn from this study. One is that there are no more than 18 subcomponents of broad factors of person description that might be replicated across languages. Saucier (1997) found that rather large attractiveness and negative valence (cf. Tellegen, 1993) factors, beyond the recognizable Big Five, can be found when a wider range of descriptors are included. Both German and English clearly contain a number of adjectives that might fall into either domain, but neither was substantially represented in the present variable selections. Moreover, Saucier and Goldberg (1998) isolated a number of clusters of personality adjectives that are outliers to the Big Five, including religiousness, folksiness, thriftiness, cunning–slyness, prejudice, and sensuality. These clusters would almost certainly be relatively independent of the present 18 but would be underrepresented here because of our focus on subcomponents of the Big Five, to which they are not related in any substantial degree. Furthermore, there may be some psychometric advantages to having an equal number of subcomponents for each factor, and therefore it might be reasonable to add more subcomponents for certain factors on nonlexical grounds. Overall, our studies have not identified the perfect set of subcomponents; rather, they have identified only some necessary features for an adequately comprehensive model of personality factor subcomponents. They must inevitably be supplemented by features from expert constructs. Indeed, arguments may eventually be made for keeping the necessary content demonstrated here but restructuring it, leading to superordinate constructs other than the Big Five.

A second inappropriate inference would be to overdramatize the magnitude of these findings from a cross-cultural perspective. German and English are rather closely related languages, having diverged less than 2,000 years ago. Different as they may be in some ways, German culture and American culture have mutually influenced one another. Germany is second only to Great Britain as a source of the immigrant ancestors of the present American population, and American culture has had a substantial impact on Germany since the Second World War. English and German are among the most widely spoken world languages; from the standpoint of cross-cultural universals, however, finding 18 personality constructs simultaneously in English and German is not as impressive as would be finding 18 simultaneous constructs in American English and, say, the language of the Kapauku in New Guinea, where a much greater cultural and historical–linguistic distance is involved. Further studies in diverse languages and cultures are needed to specify the relative generalizability of the 18 subcomponents worldwide.

For obvious reasons of scientific generalizability, it makes sense to place greater emphasis on those personality constructs that can be readily replicated across languages and cultures. But attempts to limit personality measurement to such constructs would represent another inappropriate use of the present findings. For both predictive and diagnostic purposes, some useful constructs (e.g., bulimia) may be fairly culture specific. On the other hand, identifying such culture-specific features is facilitated when one has already identified those features that readily reappear across boundaries of culture, such as, perhaps, the present subcomponents.

Summary and Conclusion

The personality terms people use—to describe both their own attributes and those of others—have a recurrent structure. Previous work has tended to assume that study of the descriptive features of the lexicon is not applicable to delineating finer grained personality subcomponents. However, the present studies show how that recurrent structure extends, to some degree, beyond broad-level factors such as the Big Five. Big Five domains, taken one at a time, were shown to have commonalities in the structure of oblique factors across two languages, English and German. We call these correlated factors "subcomponents," and the 18 detailed here provide a more finely faceted, replicated measurement model for the Big Five than has been provided by any previous study of the language of personality. If it is accepted that a broad factor is not so much one thing as a collection of many things that have something in common, the present report can serve as an initial, lower bound estimate of what those "many things" might be, for each of the Big Five. Extraversion, for example, might well be conceived as that which sociability, unrestraint, assertiveness, and active adventurousness have in common. We hope that the present article facilitates an improved critical understanding of the Big Five less subject to inaccurate generalizations.

As is true of any model based on representation in natural language, the subcomponents derived in the present study are not the last word; instead, they provide a "first word." If one wishes to specify the most important subcomponents that make up each of the Big Five personality factors, one can do far worse than to start with subcomponents, such as these, that are indifferently derivable in either of two languages. Indeed, any model or inventory that fails to incorporate their content is unlikely to be truly comprehensive.

References

Angleitner, A., Ostendorf, F., & John, O. P. (1990). Towards a taxonomy of personality descriptors in German: A psycho-lexical study. *European Journal of Personality, 4,* 89–118.

Benet-Martinez, V., & Waller, N. G. (1997). Further evidence for the cross-cultural generality of the Big Seven factor model: Indigenous and imported Spanish personality constructs. *Journal of Personality, 65,* 567–598.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117,* 187–215.

Briggs, S. R. (1989). The optimal level of measurement for personality constructs. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 246–260). New York: Springer.

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., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the*

Sixteen Personality Factor Questionnaire (16PF). Champaign, IL: Institute for Personality and Ability Testing.

Church, A. T., Reyes, J. A. S., Katigbak, M. S., & Grimm, S. D. (1997). Filipino personality structure and the Big Five model: A lexical approach. *Journal of Personality, 65,* 477–528.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12,* 887–898.

Cureton, E. E., & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika, 40,* 183–195.

De Raad, B., Hendriks, A. A. J., & Hofstee, W. K. B. (1992). Towards a refined structure of personality traits. *European Journal of Personality, 6,* 301–319.

Di Blas, L., & Forzi, M. (1998). An alternative taxonomic study of personality descriptors in the Italian language. *European Journal of Personality, 12,* 75–101.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 41, pp. 417–440). Palo Alto, CA: Annual Reviews.

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 1246–1256.

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research, 18,* 197–218.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,* 1216–1229.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–42.

Goldberg, L. R. (1993a). The structure of personality traits: Vertical and horizontal aspects. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 169–188). Washington, DC: American Psychological Association.

Goldberg, L. R. (1993b). The structure of phenotypic personality traits. *American Psychologist, 48,* 26–34.

Goldberg, L. R., & Digman, J. M. (1994). Revealing structure in the data: Principles of exploratory factor analysis. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 216–242). New York: Springer.

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology, 48,* 82–98.

Goldberg, L. R., & Saucier, G. (1995). So what do you propose we use instead? A reply to Block. *Psychological Bulletin, 117,* 221–225.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103,* 265–275.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology, 17,* 65–70.

Hofstee, W. K. B. (1990). The use of everyday personality language for scientific purposes. *European Journal of Personality, 4,* 77–88.

Hofstee, W. K. B., De Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63,* 146–163.

Hofstee, W. K. B., Kiers, H. A. L., De Raad, B., Goldberg, L. R., & Ostendorf, F. (1997). A comparison of Big-Five structures of personality traits in Dutch, English, and German. *European Journal of Personality, 11,* 15–31.

Hough, L. H., & Schneider, R. J. (in press). The frontiers of I/O personality research. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations.* San Francisco: Jossey-Bass.

Hrebickova, M., Ostendorf, F., & Angleitner, A. (1995, July). *Basic dimensions of personality description in the Czech language.* Poster session presented at the 7th meeting of the International Society for the Study of Individual Differences, Warsaw, Poland.

Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Port Huron, MI: Research Psychologists Press.

Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology, 74,* 1556–1565.

John, O. P. (1990). The 'Big Five' factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: Guilford Press.

John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology, 60,* 348–361.

Kosmitzki, C., & John, O. P. (1993). *Germans and Americans: The effects of culture on stereotypes, self-descriptions, and evaluative judgments.* Unpublished manuscript, University of California at Berkeley.

McCrae, R. R. (1990). Traits and trait names: How well is openness represented in natural languages? *European Journal of Personality, 4,* 119–129.

McCrae, R. R., & Costa, P. T. (1983). Joint factors in self-reports and ratings: Neuroticism, Extraversion, and Openness to Experience. *Personality and Individual Differences, 4,* 245–255.

McCrae, R. R., & Costa, P. T. (1985). 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. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81–90.

Mershon, B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology, 55,* 675–680.

Norman, W. T. (1967). *2800 personality trait descriptors: Normative operating characteristics for a university population.* Ann Arbor: Department of Psychology, University of Michigan.

Osgood, C. E., May, W. H., & Miron, M. S. (1975). *Cross-cultural universals of affective meaning.* Urbana: University of Illinois Press.

Ostendorf, F. (1990a). *Sprache unde Persoenlichkeitsstruktur: Zur Validitaet des Funf-Faktoren-Modells der Persoenlichkeit* [Language and personality structure: On the validity of the five-factor model of personality]. Regensburg, Germany: S. Roderer Verlag.

Ostendorf, F. (1990b). *Taxonomie einer Liste von 830 personbeschreibenden Adjektiven.* Unpublished research report, University of Bielefeld, Bielefeld, Germany.

Peabody, D. (1985). *National characteristics.* Cambridge, England: Cambridge University Press.

Peabody, D. (1987). Selecting representative trait adjectives. *Journal of Personality and Social Psychology, 52,* 59–71.

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology, 57,* 552–567.

Saucier, G. (1992a). Benchmarks: Integrating affective and interpersonal circles with the Big-Five personality factors. *Journal of Personality and Social Psychology, 62,* 1025–1035.

Saucier, G. (1992b). Openness versus intellect: Much ado about nothing? *European Journal of Personality, 6,* 381–386.

Saucier, G. (1994). Separating description and evaluation in the structure of

personality attributes. *Journal of Personality and Social Psychology, 66,* 141–154.

Saucier, G. (1997). Effects of variable selection on the factor structure of person-descriptors. *Journal of Personality and Social Psychology, 73,* 1296–1312.

Saucier, G. (1998, July). *A hierarchy of personality factors in the natural language.* Paper presented at the 9th European Conference on Personality, Guildford, United Kingdom.

Saucier, G., & Goldberg, L. R. (1996a). Evidence for the Big-Five factors in analyses of familiar English personality adjectives. *European Journal of Personality, 10,* 61–77.

Saucier, G., & Goldberg, L. R. (1996b). The language of personality: Lexical perspectives on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 21–50). New York: Guilford Press.

Saucier, G., & Goldberg, L. R. (1998). What is beyond the Big Five? *Journal of Personality, 66,* 495–524.

Shmelyov, A. G., & Pokhil'ko, V. I. (1993). A taxonomy-oriented study of Russian personality-trait names. *European Journal of Personality, 7,* 1–17.

Szarota, P. (1996). Taxonomy of the Polish personality-descriptive adjectives of the highest frequency of use. *Polish Psychological Bulletin, 27,* 343–351.

Szirmak, Z., & De Raad, B. (1994). Taxonomy and structure of Hungarian personality traits. *European Journal of Personality, 8,* 95–117.

Tellegen, A. (1993). Folk concepts and psychological concepts of personality and personality disorder. *Psychological Inquiry, 4,* 122–130.

Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C., & Costa, P. T. (1994). A description of the DSM-III-R and DSM-IV personality disorders with the five-factor model of personality. In P. T. Costa & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 41–56). Washington, DC: American Psychological Association.

Received September 29, 1997
Revision received July 23, 1998
Accepted November 23, 1998 ■