Chapter 4

The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives

Oliver P. John and Sanjay Srivastava
University of California, Berkeley

Taxonomy is always a contentious issue because the world does not come to us in neat little packages

—GOULD (1981, p. 158)

Personality has been conceptualized from a variety of theoretical perspectives, and at various levels of abstraction or breadth (John, Hampson, & Goldberg, 1991; McAdams, 1995). Each of these levels has made unique contributions to our understanding of individual differences in behavior and experience. One frequently studied level is personality traits (John & Gosling, in press). However, the number of personality traits, and scales designed to measure them, has escalated without an end in sight (Goldberg, 1971). Researchers, as well as practitioners in the field of personality assessment, have been faced with a bewildering array of personality scales from which to choose, with little guidance and no overall rationale at hand. To make matters worse, scales with the same name often measure concepts that are not the same, and scales with different names often measure concepts that are quite similar. Although diversity and scientific pluralism are useful, the systematic accumulation of findings and the communication among researchers has become difficult amidst the Babel of concepts and scales.

Many personality researchers had hoped that they might devise the structure that would transform the Babel into a community speaking a common language. However, such an integration was not to be achieved by any one researcher or by any one theoretical perspective. As Allport once put it, "each assessor has his own pet units and uses a pet battery of diagnostic devices" (1958, p. 258).

What personality psychology needed was a descriptive model, or taxonomy, of traits. One of the central goals of scientific taxonomies is the definition of overarching domains within which large numbers of specific instances can be understood in a simplified way. Thus, in personality psychology, a taxonomy would permit researchers to study specified domains of personality characteristics, rather than examining separately the thousands of particular attributes that make each human being individual and unique. Moreover, a generally accepted taxonomy would greatly facilitate the accumulation and communication of empirical findings by offering a standard vocabulary, or nomenclature.

After decades of research, the field is finally approaching consensus on a general taxonomy of personality traits, the "Big Five" personality dimensions. These dimensions do not represent a particular theoretical perspective but were derived from analyses of the natural language terms people use to describe themselves and others. Rather than replacing all previous systems, the Big Five taxonomy serves an integrative function because it can represent diverse systems of personality description in a common framework. It thus provides a starting place for vigorous research and theorizing that can eventually lead to an explication and revision of the descriptive taxonomy in causal and dynamic terms.

In this chapter, we first review the history of the Big Five, including the discovery of the five dimensions, research replicating and extending the model, its convergence with research in the questionnaire tradition, and the development of several instruments to measure the Big Five. Then, we compare three of the most frequently used instruments and report data regarding their reliability and convergent validity. Finally, we address a number of critical issues, including how the Big Five taxonomy is structured hierarchically, whether the five dimensions predict important life outcomes, how they develop, how they combine into personality types, and whether they are descriptive or explanatory concepts.

THE LEXICAL APPROACH AND DISCOVERY OF THE BIG FIVE

One starting place for a shared taxonomy is the natural language of personality description. Beginning with Klages (1926), Baumgarten (1933), and Allport and Odbert (1936), various psychologists have turned to the natural language as a source of attributes for a scientific taxonomy. This work, beginning with the extraction of all personality-relevant terms from the dictionary, has generally been guided by the lexical hypothesis (see John, Angleitner, & Ostendorf, 1988; Saucier & Goldberg, 1996b). The lexical hypothesis posits that most of the socially relevant and salient personality characteristics have become encoded in the natural language (e.g., Allport, 1937). Thus, the personality vocabulary contained in the dictionaries of a natural language provides an extensive, yet finite, set of attributes that the people speaking that language have found important and useful in their daily interactions (Goldberg, 1981).

Allport and Odbert's Psycholexical Study: Traits, States, Activities, and Evaluations

Following Baumgarten's (1933) work in German, Allport and Odbert (1936) conducted a seminal lexical study of the personality-relevant terms in an unabridged English dictionary. They included all the terms that could be used to "distinguish the behavior of one human being from that of another" (p. 24) Their complete list amounted to almost 18,000 terms. At the time, the staggering size of this list seemed "like a semantic nightmare" (Allport, 1937, pp. 353–354). Allport and Odbert thought that organizing these thousands of personality attributes into a satisfactory taxonomy would keep psychologists "at work for a life time" (1936, p. vi). Indeed, this task has occupied personality psychologists for more than 60 years. (For detailed reviews of the history of the lexical approach, see John et al., 1988; John, 1990).

Allport and Odbert (1936) tried to bring some order to the semantic nightmare they had created. What kinds of person descriptors are included in the dictionary? Allport and Odbert identified four major categories. The first category included personality traits (e.g., sociable, aggressive, and fearful), which they defined as "generalized and personalized determining tendencies—consistent and stable modes of an individual's adjustment to his environment" (p. 26). The second category included temporary states, moods, and activities, such as afraid, rejoicing, and elated. The third category consisted of highly evaluative judgments of personal conduct and reputation, such as excellent, worthy, average, and irritating. Although these terms presuppose some traits within the individual, they do not indicate the specific attributes that gave rise to the individual's evaluation by others or by society in general. The last category included physical characteristics, capacities and talents, terms of doubtful relevance to personality, and terms that could not be assigned to any of the other three categories.

Norman (1967) subsequently elaborated Allport and Odbert's initial classification and divided the domain into seven content categories: stable "biophysical" traits, temporary states, activities, social roles, social effects, evaluative terms, anatomical and physical terms, as well as ambiguous and obscure terms not considered useful for personality descriptive purposes. These categories illustrate that the personality

lexicon in the natural language includes a wealth of concepts. Individuals can be described by their enduring *traits* (e.g., irrascible), by the *internal states* they typically experience (furious), by the *physical states* they endure (trembling), by the *activities* they engage in (screaming), by the *effects* they have on others (frightening), by the *roles* they play (murderer), and by social *evaluations* of their conduct (unacceptable, bad). Moreover, individuals differ in their anatomical and morphological characteristics (short) and in the personal and societal evaluations attached to these appearance characteristics (cute).

Both Allport and Odbert (1936) and Norman (1967) classified the terms culled from the dictionary into mutually exclusive categories. An inspection of the classifications quickly shows that the categories overlap and have fuzzy boundaries, leading some researchers to conclude that distinctions between classes of personality descriptors are arbitrary and should be abolished (Allen & Potkay, 1981). In contrast, Chaplin, John, and Goldberg (1988) argued for a prototype conception in which each category is defined in terms of its clear cases rather than its boundaries; category membership need not be discrete but can be defined as continuous. Chaplin and colleagues (1988) applied this prototype conception to traits, states, and activities. Although the classification of a few descriptors was difficult, the core of each category was distinct from the others and could be differentiated by a set of conceptually derived attributes. Prototypical *states* were seen as temporary, brief, and externally caused. Prototypical *traits* were seen as stable, long lasting, and internally caused, and needed to be observed more frequently and across a wider range of situations than states before they were attributed to an individual. These findings closely replicated the earlier classifications and confirmed that the conceptual definitions of traits and states are widely shared.

Identifying the Major Dimensions of Personality Description: Cattell's Early Efforts

Allport and Odbert's (1936) classifications provided some initial structure for the personality lexicon. However, to be of practical value, a taxonomy must provide a systematic framework for distinguishing, ordering, and naming individual differences in people's behavior and experience (John, 1989). Aiming for such a taxonomy, Cattell (1943) used the Allport and Odbert list as a starting point for his multidimensional model of personality structure. Because the size of that list was too overwhelming for research purposes, Cattell (1943, 1945a, 1945b) began with the subset of 4,500 trait terms. Most taxonomic research has focused on the personality trait category, although the other categories are no less important. For example, the emotional-state and social-evaluation categories have recently received considerable attention (Almagor, Tellegen & Waller, 1995; Benet-Martinez & Waller, 1997).

Using both semantic and empirical clustering procedures as well as his own reviews of the personological literature available at the time (for reviews, see John et al., 1988; John, 1990), Cattell reduced the 4,500 trait terms to a mere 35 variables. That is, Cattell eliminated more than 99% of the terms Allport (1937) had so tenaciously defended. This drastic reduction was dictated primarily by the data-analytic limitations of his time, which made factor analyses of large variable sets prohibitively costly and complex. Using this small set of variables, Cattell conducted several oblique factor analyses and concluded that he had identified 12 personality factors, which eventually became part of his 16 Personality Factors (16PF) questionnaire (Cattell, Eber, & Tatsuoka, 1970).

Cattell also claimed that his factors showed excellent correspondence across methods, such as self-reports, ratings by others, and objective tests; however, these claims have not gone unquestioned (e.g., Becker, 1960; Nowakowska, 1973). Moreover, reanalyses of Cattell's own correlation matrices by others have not confirmed the number and nature of the factors he proposed (e.g., Tupes & Christal, 1961[1]). Digman and Takemoto-Chock (1981) concluded that Cattell's "original model, based on the unfortunate clerical errors noted here, cannot have been correct" (p. 168), although the second-order factors of the 16PF show some correspondence between Cattell's system and the subsequently derived Big Five dimensions.

THE BIG FIVE FACTORS IN PERSONALITY TRAIT RATINGS

Discovery of the Big Five in Cattell's Variable List

Cattell's pioneering work, and the availability of a relatively short list of variables, stimulated other researchers to examine the dimensional structure of trait ratings. Several investigators

were involved in the discovery and clarification of the Big Five dimensions. Fiske (1949) constructed much simplified descriptions from 22 of Cattell's variables; the factor structures derived from self-ratings, ratings by peers, and ratings by psychological staff members were highly similar and resembled what would be later known as the Big Five. To clarify these factors, Tupes and Christal (1961) reanalyzed correlation matrices from eight different samples, ranging from airmen with no more than high school education to first-year graduate students, and included ratings by peers, supervisors, teachers, or experienced clinicians in settings as diverse as military training courses and sorority houses. In all the analyses, Tupes and Christal found "five relatively strong and recurrent factors and nothing more of any consequence" (1961, p. 14).

This five-factor structure has been replicated by Norman (1963), Borgatta (1964), and Digman and Takemoto-Chock (1981) in lists derived from Cattell's 35 variables. These factors are typically labeled:

I. Extraversion or Surgency (talkative, assertive, energetic)
II. Agreeableness (good-natured, cooperative, trustful)
III. Conscientiousness (orderly, responsible, dependable)
IV. Emotional Stability versus Neuroticism (calm, not neurotic, not easily upset)
V. Intellect or Openness (intellectual, imaginative, independent-minded)

These factors eventually became known as the "Big Five" (Goldberg, 1981)—a title chosen not to reflect their intrinsic greatness but to emphasize that each of these factors is extremely broad. Thus, the Big Five structure does not imply that personality differences can be reduced to only five traits. Rather, these five dimensions represent personality at the broadest level of abstraction, and each dimension summarizes a large number of distinct, more specific personality characteristics.

Testing the Big Five in a Comprehensive Set of English Trait Terms

After a period of dormancy during the 1970s and early 1980s, research on the Big Five, and on issues of personality structure more generally, has increased dramatically since the mid-1980s. Factor structures resembling the Big Five were identified in numerous sets of variables (e.g., Botwin & Buss, 1989; Conley, 1985; De Raad, Mulder, Kloosterman, & Hofstee, 1988; Digman & Inouye, 1986; Field & Millsap, 1991; Goldberg, 1981, 1990; John, 1990; McCrae & Costa, 1985c, 1987; Peabody & Goldberg, 1989; Saucier & Goldberg, 1996a). However, a number of these studies were influenced by Cattell's selection of variables (Block, 1995), making it important to test the comprehensiveness and generality of the Big Five in more comprehensive variable sets. To update the Allport and Odbert list and to rectify the imperfections of Cattell's reduction steps, Norman (1967) compiled an exhaustive list of personality descriptive terms, which he sorted into 75 semantic categories. Goldberg (1990; see also 1981, 1982) used this list to clarify the nature and composition of these broad factors and to test their stability and generalizability across methodological variations and data sources. Using Norman's (1967) listing, Goldberg (1990) constructed an inventory of 1,710 trait adjectives that participants could use to rate their own personality. He then scored Norman's semantic categories as scales and factor analyzed their intercorrelations in the self-rating data. The first five factors represented the Big Five and replicated across a variety of different methods of factor extraction and rotation. Moreover, Goldberg (1990) demonstrated that the first five factors remained virtually invariant when more than five were rotated.

To ensure independence from any a priori classification, Goldberg (1990) conducted two additional studies using abbreviated sets of more common terms. In one study, Goldberg obtained self and peer ratings of 475 very common trait adjectives which he had grouped into 131 sets of "tight synonym" clusters. In four samples, the five-factor structures were very similar to each other and to the structure obtained in the more comprehensive list of 1,710 terms, and the results in the self-rating data were virtually indistinguishable from those in the peer ratings. Most important, however, were the results from the search for replicable additional factors. In a more recent study, Saucier and Goldberg (1996a) selected 435 trait adjectives rated by subjects as highly familiar terms; a factor analysis of these adjectives closely replicated the Big Five. Furthermore, a thorough search for factors beyond the Big Five showed that the Big Five

were the only consistently replicable factors (Saucier, 1997).

Assessing the Big Five with Trait Descriptive Adjectives

Goldberg (1990, 1992) distilled his extensive taxonomic findings into several published adjective lists. One of them is a 50-item instrument using the so-called "transparent format" (Goldberg, 1992), which is excellent for instructional purposes (Pervin & John, 1997). For each factor, this measure presents 10 bipolar adjective scales (e.g., quiet–talkative) grouped together under the factor name, thus making the constructs being measured transparent to the research participants. The list used more commonly in research is the set of 100 unipolar trait descriptive adjectives (TDA). Goldberg (1992) conducted a series of factor analytic studies to develop and refine the TDA as an optimal representation of the five-factor space in English, selecting for each Big Five scale only those adjectives that uniquely defined that factor. These scales have impressively high internal consistency, and their factor structure is easily replicated.[2]

Another adjectival measure of the Big Five was developed by Wiggins (1995; Trapnell & Wiggins, 1990). In his 20-year program of research on the interpersonal circumplex, Wiggins (1979) has used personality trait adjectives to elaborate both the conception and the measurement of the two major dimensions of interpersonal behavior, Dominance (or Agency) and Nurturance (or Communion). Noting that the first dimension closely resembles the Extraversion factor in the Big Five, and the second dimension the Agreeableness factor, Wiggins extended his circumplex scales by adding adjective measures for the other three of the Big Five factors (Trapnell & Wiggins, 1990). The resulting Interpersonal Adjective Scales (Wiggins, 1995) have excellent reliabilities and converge well with other measures; they have been used by researchers who want to measure the specific octants of the interpersonal circle as well as the Big Five.

The circumplex approach has also been applied to a perennial problem in lexical research on personality factors. One important task is to spell out, with much more precision, those characteristics that fall in the fuzzy regions *between* the factors. Using 10 two-dimensional circumplexes, Hofstee, De Raad, and Goldberg (1992) have devised a novel empirical approach to represent the space formed by each pair of factors. This approach specifies facets that reflect various combinations of two factors. The facets differ in whether they are more closely related to one or the other factor. For example, there are two facets that reflect high Agreeableness and high Conscientiousness, but they differ in which of the two factors is given prominence. Thus, the *responsibility* facet represents agreeable Conscientiousness, whereas the *cooperation* facet represents conscientious Agreeableness (Hofstee, Kiers, De Raad, Goldberg, & Ostendorf, 1997).

Cross-Language and Cross-Cultural Studies

The results reviewed so far suggest that the Big Five structure provides a replicable representation of the major dimensions of trait description in English. The five-factor structure seems to generalize reliably across different types of samples, raters, and methodological variations when comprehensive sets of variables are factored. Generalizability across languages and cultures is another important criterion for evaluating personality taxonomies (John, Goldberg, & Angleitner, 1984).

Taxonomic research in other languages and cultures can determine the usefulness of a taxonomy across cultural contexts and test for universals and variations in the encoding of individual differences across languages and cultures (Goldberg, 1981). The existence of cultural universals would be consistent with an evolutionary interpretation of the way individual differences have become encoded as personality categories into the natural language: If the tasks most central to human survival are universal, the most important individual differences, and the terms people use to label these individual differences, would be universal as well (Buss, 1996; Hogan, 1983; see also Buss, Chapter 2, this volume). Similarly, if cross-cultural research reveals a culturally specific dimension, variation on that dimension may be uniquely important within the particular social context of that culture (Yang & Bond, 1990).

Although central from the vantage point of the lexical approach, cross-language research is difficult and expensive to conduct, and until the 1990s it was quite rare. In the initial comprehensive taxonomic studies, English was the language of choice, primarily because the taxonomers were American (for reviews, see John et al., 1984; John et al., 1988).

Initial Studies in Dutch and German

The first two non-English taxonomy projects involved Dutch and German, languages closely related to English. The Dutch project has been carried out by Hofstee, De Raad, and their colleagues at the University of Groningen in the Netherlands (De Raad, Mulder, Kloosterman, & Hofstee, 1988; Hofstee et al., 1997; see also De Raad, Perugini, et al., 1998, for reviews). The conclusions from the Dutch projects are generally consistent with those from the American English research: Only five factors were replicable across different selections of trait adjectives and across different subject samples. Those five factors were similar to the English Big Five, although in Dutch the fifth factor emphasizes Unconventionality and Rebelliousness rather than Intellect and Imagination as found in English.

The dictionary-based German taxonomy project was begun in Bielefeld by Angleitner, Ostendorf, and John (1990), who carried out a "psycholexical" study of the German personality vocabulary. Their study was explicitly based on the prototype conception and improved on the earlier studies of English in several respects. In particular, 10 independent judges classified all the terms, thus providing a continuous measure of prototypicality and an assessment of the reliability and validity of the judgments. The resulting German personality lexicon is more convenient to use than the unwieldy Allport and Odbert lists because continuous prototypicality values are available for each term in 13 different content categories. Thus, it is easy to select subsets of prototypical traits, states, social evaluations, and so on from the total pool for further studies. Angleitner and colleagues' (1990) research served as a blueprint for several taxonomic efforts in other languages.

Ostendorf (1990) selected the most prototypical trait adjectives from the German taxonomy, and his factor analyses of about 450 traits yielded the clearest replication of the Big Five so far. In addition to the prototypical traits representing the distillation of the German trait lexicon, Ostendorf also included German translations of several English Big Five instruments. Thus, Ostendorf's study is a good example of the combined emic–etic design, which allows researchers to establish empirically the similarity of indigenous (emic) factors to the factors translated from other languages and cultures (etic). Correlational analyses allowed Ostendorf to demonstrate substantial convergence between the emic German dimensions and the etic Big Five measures in the same sample of German subjects.

However, this combined emic–etic strategy is difficult to implement and not consistently used in research. Thus, conclusions about factor similarity are often made by "eyeballing" the item content of the factors in the indigenous language and comparing it to the typical factor definitions in English. That leaves much leeway to the investigators in "seeing" a factor that another investigator might not see. For example, the Hebrew factor defined primarily by traits such as sophisticated, sharp, knowledgeable, articulate, and impressive would lead some researchers to see an Intellect factor, whereas Almagor and colleagues (1995) interpreted it as Positive Valence.

Underestimating Cross-Language Congruence

One of the difficulties in cross-language research involves translations. Often, researchers working within their indigenous language have to translate their concepts into English to communicate their findings and much slippage occurs in the translation process. For example, one wonders why "temperamental" was a definer of Extraversion in German until one realizes that the German trait was probably *temperamentvoll*, which has nothing to do with temper but means "full of life and energy," as in vivacious. Similarly, *frizzante* (translated as sparkling) was not related to brilliant intellect, but instead seems to mean something like the English word "bubbly."

An initial study of German–English bilinguals, which provided support for cross-language generalizability (John et al., 1984), directly addressed the issue of translation equivalence. The unique advantage of the bilingual design is that sample differences can be controlled and that translation checks can be made at the level of individual items because the same subject provides descriptions in both languages (see also Benet-Martinez & John, 1998). Using a careful back-translation procedure, John and colleagues (1984) found acceptable levels of translation equivalence between English and German trait adjectives, with a mean correlation of .52 across a two-week interval between administrations. However, several translations proved to be inadequate, with item-translation correlations approaching zero. These findings suggest that mistranslations that cannot be

detected in monolingual investigations will lead to severe underestimations of cross-language generality.

To permit empirical estimates of factor similarity across monolingual investigations, Hofstee, De Raad, and their colleagues have used translations of terms as a way to compare factor solutions across languages. For example, Hofstee and colleagues (1997) identified 126 words that they could translate across previous lexical studies in English, Dutch, and German and used them to assess factor congruence coefficients among all pairs of factors in the three languages. Their findings are illuminating in that they showed considerable congruence across these three Germanic languages. With the exception of the fifth factor in Dutch and English, the pairwise congruence coefficients all exceeded .70. Strangely, the authors interpreted these levels of cross-language congruence as "disappointing" (Hofstee et al., 1997, p. 27). This interpretation contradicts Ostendorf's (1990) own conclusions, which were drawn from the emic–etic comparisons in his well-designed study.

We are more optimistic about these findings. The empirically observed levels of factor congruence reported by Hofstee and colleagues (1997) can be interpreted only if one assumes that the translations are perfectly equivalent and that the factor structures in each language are perfectly stable. What happens when we correct the cross-language congruence coefficients at least for the imperfect reliability of the factor structures reported by Hofstee and colleagues? The corrected English–German congruence coefficients range from .84 to .93, impressive values given that they are not corrected for the imperfect translations; moreover, the correspondence for the fifth factor was .93, suggesting that the Intellect/Openness factor was defined almost identically in English and German. The corrected English–Dutch and German–Dutch congruence coefficients were very similar to each other, and suggested the same conclusions: Congruence was substantial for the first four factors (.88 to .97) but not the fifth (.50 to .53). In short, our reexamination suggests that translation-based comparisons across languages are heuristically useful but should not be interpreted in terms of absolute effect sizes. These results also suggest that the fifth factor in Dutch is defined differently in the other two languages, and explanations for this finding need to be sought.

Rules for Including Trait Descriptors in Taxonomic Studies

In all likelihood, some of the differences observed among the factor structures in the three languages also result from the different inclusion rules followed by the taxonomy teams. The selection criterion used by the Dutch researchers favored terms related to temperament, excluded terms related to intellect, talents, and capacities, and included a number of extremely negative evaluative terms, such as perverse, sadistic, and criminal. The German team explicitly included intellect and talent descriptors but omitted attitudes and evaluative terms, which were included as categories separate from traits. Finally, the American English taxonomy included attitudinal terms such as liberal, progressive, and provincial, along with a number of intellect terms. Given the diverse range of traits related to the fifth factor, it is less surprising that the German and English factors shared the intellect components, whereas the Dutch factor included some imagination-related traits (e.g., inventive, original, imaginative) but otherwise emphasized unconventionality and was thus interpreted initially as a "Rebelliousness" factor. An Italian taxonomy (Caprara & Perugini, 1994) found a similar fifth factor interpreted as Unconventionality: Not surprisingly, these Italian researchers had followed the Dutch selection procedures rather than the German procedures, which would have represented more Intellect terms in the taxonomy.

Szirmak and De Raad (1994) examined Hungarian personality descriptors and found strong support for the first four of the Big Five but failed to obtain a factor resembling the fifth of the Big Five; instead, when they forced a five-factor solution, the Agreeableness factor split into two factors. An Intellect/Openness factor emerged only when six factors were rotated. Again, this finding may be due to the selection rules that included a "trait versus state rating."

Evidence in Non-Germanic Languages

The personality lexicon has recently been studied in a wide range of additional languages, such as Chinese (Yang & Bond, 1990), Czech (Hrebickova & Ostendorf, 1995), Hebrew (Almagor et al., 1995), Hungarian (Szirmak & De Raad, 1994), Italian (De Raad, Di Blas, & Perugini, 1998), Polish (Szarota, 1995), Russian (Shmelyov & Pokhilko, 1993), and Turkish (Somer &

Goldberg, 1999). A recent review (De Raad, Perugini, et al., 1998) has compared many of the European studies, using translations to estimate factor similarity quantitatively. Most generally, factors similar to the Big Five have been found in many other languages but often, more than five factors needed to be rotated and sometimes two indigenous factors corresponded to one of the Big Five. Overall, the evidence is least compelling for the fifth factor, which appears in various guises, ranging from pure Intellect (in German) to Unconventionality and Rebelliousness (in Dutch and Italian).

Extensions into cultures different from the industrialized West have also begun to appear. Whereas early studies used translations of English-language measures (Bond, 1979, 1983; Bond & Forgas, 1984; Bond, Nakazato, & Shiraishi, 1975; Guthrie & Bennett, 1971; Nakazato, Bond, & Shiraishi, 1976; White, 1980), more recent studies have used emic and combined emic–etic designs. For example, extensive studies of Filipino samples have provided some support for the generality of the Big Five (Church & Katigbak, 1989; Church, Reyes, Katigbak, & Grimm, 1997). Church and Katigbak (1989) had subjects generate behavioral exemplars, and Church and colleagues (1997) derived a comprehensive list of personality descriptors following the methods proposed by the German taxonomy team. Both studies suggest that the structure of the Filipino personality lexicon is quite similar to the Big Five, although more than five factors needed to be extracted to produce all of the Big Five dimensions. As the authors caution, "this does not mean that there are no unique concepts in either language. However, at a higher level of generality, similar structural dimensions emerge" (Church & Katigbak, 1989, p. 868).

Bond and collaborators (Yang & Bond, 1990; Yik & Bond, 1993) have recently followed up on their earlier etic work in Chinese. They drew their emic items from free descriptions and from indigenous personality questionnaires. By including translations of Big Five marker items from English, they were able to use regression analyses to compare the emic factor space with the etic (i.e., imported) Big Five. Their results suggest that although the Chinese language does not cleanly reproduce the English Big Five and several differences remain, the indigenous Chinese dimensions do overlap considerably with the Big Five dimensions.

In summary, the cross-language research suggests that the Big Five can be replicated in Germanic languages. The evidence for non-Western languages and cultures is more complex, and Factor V generally shows the weakest replicability. Thus, strong conclusions about the linguistic universality of the lexically derived Big Five would be premature. Most generally, we agree with De Raad, Perugini, and colleagues (1998) who concluded that the findings in seven languages support "the general contours of the Big Five model as the best working hypothesis of an omnipresent trait structure" (p. 214).

THE BIG FIVE IN PERSONALITY QUESTIONNAIRES

While researchers in the lexical tradition were accumulating evidence for the Big Five, the need for an integrative framework became more pressing among researchers who studied personality with questionnaire scales. Joint factor analyses of questionnaires developed by different investigators had shown that two broad dimensions, Extraversion and Neuroticism, appear in one form or another in most personality inventories. Beyond these "Big Two" (Wiggins, 1968), however, the various questionnaire-based models had shown few signs of convergence. For example, Eysenck (1991) observed that "Where we have literally hundreds of inventories incorporating thousands of traits, largely overlapping but also containing specific variance, each empirical finding is strictly speaking only relevant to a specific trait. . . . This is not the way to build a unified scientific discipline" (p. 786).

Costa and McCrae's Research

The situation began to change in the early 1980s when Costa and McCrae were developing the NEO Personality Inventory (eventually published in 1985) to measure three broad personality dimensions: Neuroticism, Extraversion, and Openness to Experience. Costa and McCrae (1976) had begun their work with cluster analyses of the 16PF (Cattell, Eber, & Tatsuoka, 1970) which, as we described above, originated in Cattell's early lexical work. Their analyses again yielded the ubiquitous Extraversion and Neuroticism dimensions, but also convinced Costa and McCrae of the importance of Openness, which originated from several of Cattell's primary factors (e.g., imaginative, experimenting).

In 1983 Costa and McCrae realized that their NEO system closely resembled three of the Big Five factors, but did not encompass traits in the Agreeableness and Conscientiousness domains. They therefore extended their model with preliminary scales measuring Agreeableness and Conscientiousness. In several studies, McCrae and Costa (1985b, 1985c, 1987) demonstrated that their five questionnaire scales converged with adjective-based measures of the Big Five, although their conception of Openness seemed broader than the Intellect or Imagination factor emerging from the lexical analyses (Saucier & Goldberg, 1996a). A series of influential papers showed that these five factors could also be recovered in various other personality questionnaires, as well as in self-ratings on Block's (1961/1978) California Adult Q-set (see Costa & McCrae, 1992; McCrae & Costa, 1990).

The Revised NEO Personality Inventory

The initial NEO Personality Inventory (Costa & McCrae, 1985) included scales to measure six facets of Neuroticism, Extraversion, and Openness but did not include any facet scales for the newly added Agreeableness and Conscientiousness. In 1992, Costa and McCrae published the 240-item NEO Personality Inventory, Revised (NEO PI-R; Costa & McCrae, 1992), which permits differentiated measurement of each Big Five dimension in terms of six specific facets per factor (Costa & McCrae, 1995). Table 4.1 shows the six facets defining each of the factors, as well as a highly correlated trait adjective to illustrate the links with the lexical research. The NEO PI-R was developed in samples of middle-aged and older adults, using both factor analytic and multimethod validational procedures of test construction. The scales have shown substantial in-

TABLE 4.1. Costa and McCrae's (1992) NEO PI-R Facets

Big Five dimensions		Facet (and correlated trait adjective)[a]
E	Extraversion versus introversion	Gregariousness (sociable)
		Assertiveness (forceful)
		Activity (energetic)
		Excitement-seeking (adventurous)
		Positive emotions (enthusiastic)
		Warmth (outgoing)
A	Agreeableness versus antagonism	Trust (forgiving)
		Straightforwardness (not demanding)
		Altruism (warm)
		Compliance (not stubborn)
		Modesty (not show-off)
		Tender-mindedness (sympathetic)
C	Conscientiousness versus lack of direction	Competence (efficient)
		Order (organized)
		Dutifulness (not careless)
		Achievement striving (thorough)
		Self-discipline (not lazy)
		Deliberation (not impulsive)
N	Neuroticism versus emotional stability	Anxiety (tense)
		Angry hostility (irritable)
		Depression (not contented)
		Self-consciousness (shy)
		Impulsiveness (moody)
		Vulnerability (not self-confident)
O	Openness versus closedness to experience	Ideas (curious)
		Fantasy (imaginative)
		Aesthetics (artistic)
		Actions (wide interests)
		Feelings (excitable)
		Values (unconventional)

[a]These traits from the Adjective Check List (listed in parentheses following each facet) correlated substantially with scores on that facet in a study of self-ratings (Costa & McCrae, 1992, p. 49).

ternal consistency, temporal stability, and convergent and discriminant validity against spouse and peer ratings (Costa & McCrae, 1992; McCrae & Costa, 1990). Moreover, the factor structure of the 30-facet scales replicates very closely in a broad range of languages and cultures (McCrae & Costa, 1997).

For many research applications, the NEO PI-R is rather lengthy. To provide a shorter measure, Costa and McCrae (1992) developed the 60-item NEO-FFI, an abbreviated version based on an item factor analysis of the 1985 version of the NEO PI (Costa & McCrae, 1985). The 12-item scales of the FFI include the items that loaded highest on each of the five factors in that analysis. The item content of the scales was adjusted somewhat to ensure adequate content coverage of the facets; however, these scales do not equally represent each of the six facets defining each factor. For example, the Agreeableness scale includes five items from the Altruism facet, three from Compliance, two from Trust, one from Tender-Mindedness, one from Straightforwardness, and none from Modesty. The reliabilities reported in the manual (Costa & McCrae, 1992) are adequate, with a mean of .78 across the five scales. The NEO-FFI scales are substantially correlated with the NEO PI-R scales, suggesting that they inherit a substantial portion of the validity of the longer scales.

A PROTOTYPE APPROACH TO DEFINING THE BIG FIVE ACROSS STUDIES

So far, we have reviewed both Goldberg's (1990) lexically based research and Costa and McCrae's (1992) questionnaire-based research on the Big Five. Despite these extensive studies, the Big Five structure has not been accepted as a taxonomic superstructure by all researchers in the field (e.g., Block, 1995; Eysenck, 1992, 1997; McAdams, 1992; Pervin, 1994). One problem, it seems, is the perception that there is no *single* Big Five, which is evident in questions such as "*which* Big Five?" or "*whose* Big Five?" (John, 1989). For example, across studies the Extraversion factor has appeared as confident self-expression, surgency, assertiveness, social extraversion, and power (see John, 1990, Table 3.1). Agreeableness has been labeled social adaptability, likability, friendly compliance, agreeableness, and love. The Conscientiousness factor has appeared under the names dependability, task interest, will to achieve, impulse control, and work. Neuroticism versus Emotional Stability has also been called emotionality, ego strength (vs. anxiety), dominant–assured, satisfaction, and affect. Finally, Openness has also been labeled inquiring intellect, culture, intelligence, intellect, intellectual interests, and intellectance.

Of course, some variation from study to study is to be expected with dimensions as broad and inclusive as the Big Five. Differences in factor solutions are likely to arise when researchers differ in the variables they include, thus representing different parts of the factor's total range of meaning. Moreover, researchers differ in their preferences for factor labels even when the factor content is quite similar. The fact that the labels differ does not necessarily mean that the factors are different, too. Thus, there may be more commonality than meets the eye.

A prototype approach may help identify these commonalities across studies. Natural categories typically have fuzzy and partially overlapping definitions (Rosch, 1978), and the Big Five are no exception. Fuzzy categories may still be useful if they can be defined in terms of prototypical exemplars. Similarly, the Big Five may be defined with prototypical traits that occur consistently across studies.

How might one integrate the findings from a large and varied set of factor analytic investigations, each using somewhat different sets of variables, analytic procedures, and factor interpretations? One approach is to conceptually map the five dimensions into a common language. To abstract the common elements in these findings, John (1989, 1990) used human judges, and the 300 terms included in the Adjective Check List (ACL; Gough & Heilbrun, 1983) served as the standard language.

Conceptually Derived Prototype Descriptions of the Big Five

A set of 10 judges first formed a detailed understanding of the Big Five dimensions by reviewing the factor solutions and interpretations of all the important articles published on the Big Five by that time. The judges then independently sorted each of the 300 items in the ACL into one of the Big Five domains, or, if that was not possible, into a sixth "other" category. Interjudge agreement was substantial; coefficient alpha reliabilities ranged from .90 for Factor IV to .94 for Factor V, suggesting that the raters had formed a consensually shared understanding of the five di-

mensions. As shown in Table 4.2, 112 of the 300 ACL terms were assigned to one of the Big Five with almost perfect agreement (i.e., by at least 90% of the judges). These terms form a relatively narrow, or "core," definition of the five factors because they include only those traits that appeared consistently across studies.

As with any rationally constructed measure, the validity of these categorizations must be tested empirically. The results from a factor analysis of the 112 terms are also included in Table 4.2. If the initial prototypes adequately capture the composition of the Big Five, the 112 terms should clearly define five factors, and each term should load only on its respective factor. Most research on the Big Five has been based on self- and peer ratings, typically by college students. This study used judgments by psychologists based on intensive observations and interviews, thus testing the degree to which the Big Five can capture the personality judgments formulated by psychologists and addressing the criticism that the Big Five merely captures the personality conceptions of lay persons (Block, 1995).

Validation of the Prototypes in Observer Data

The ACL was initially developed at the Institute of Personality Assessment and Research (now the Institute of Personality and Social Research, or IPSR) in Berkeley, California, as a procedure to help staff members describe the personalities of individuals examined in assessment programs (Gough & Heilbrun, 1983, p. 1). John (1990) used a sample of 140 men and 140 women who had participated in groups of 10 to 15 in one of the IPSR assessment weekends. As each subject had been described on the ACL by 10 staff members, a factor analysis using these aggregated observer judgments could be performed. The varimax rotated factor loadings, shown in Table 4.2 for each adjective for its hypothesized factor, provide a compelling confirmation of the initial prototypes. With one exception, each item loaded on its hypothesized factor in the expected direction; for 98 of the 112 items the highest loading was also on that factor, and most of the loadings were substantial.

Note that the items defining each of the factors cover a broad range of content. For example, the Extraversion factor includes traits such as active, adventurous, assertive, dominant, energetic, enthusiastic, outgoing, sociable, and show-off. In light of the enormous breadth of the five factors, the heterogeneity of the previous factor labels is more easily understood. Different investigators have focused on different components, or facets, of the total range of meaning subsumed by each factor. In this study, the Extraversion factor includes at least five distinguishable components: Activity level (active, energetic), Dominance (assertive, forceful, bossy), Sociability (outgoing, sociable, talkative), Expressiveness (adventurous, outspoken, noisy, show-off), and Positive emotionality (enthusiastic, spunky). Note that these five components are similar to five of the six facets Costa and McCrae (1992) included in their definition of the Extraversion domain—Activity, Assertiveness, Gregariousness, Excitement-seeking, and Positive Emotions. Their sixth facet, Warmth, is here considered a component of Agreeableness (Factor II); all 10 judges interpreted past research to imply that Warmth is part of Factor II, and the empirical loading of .82 confirmed this interpretation. In addition to Warmth (affectionate, gentle, warm), Factor II covers themes such as Tender-Mindedness (sensitive, kind, soft-hearted, sympathetic), Altruism (generous, helping, praising), and Trust (trusting, forgiving), as contrasted with Hostility, Criticality, and Distrust; again, note the convergence with Costa and McCrae's (1992) facets. More generally, the definitions of the Big Five in Table 4.2 seem to capture the prototypical traits found in other studies.

The Prototypical Definition of Factor V: Culture, Intellect, or Openness?

The findings in Table 4.2 also address a recurrent issue in the literature, namely, how the fifth factor should be defined. Most of the deviations from the hypothesized structure (marked by asterisks in Table 4.2), involved Factor V. Many items referring to aspects of culture (i.e., civilized, polished, dignified, foresighted, logical) loaded more highly on Factor III (Conscientiousness) than on Factor V, thus further discrediting a Culture interpretation of Factor V. The items that did load substantially on the fifth factor include both the "open" characteristics (e.g., artistic, curious, original, wide interests) highlighted by McCrae and Costa (1985b, 1985c) and the "intellectual" characteristics (intelligent, insightful, sophisticated) emphasized by Digman and Inouye (1981), Peabody and Goldberg (1989), and Goldberg (1990).

How do these findings compare with other research? Goldberg's (1990, 1992) detailed lexical

TABLE 4.2. Initial and Validated Big Five Prototypes: Consensually Selected ACL Marker Items and Their Factor Loadings in Personality Descriptions Obtained from 10 Psychologists Serving as Observers

Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness/Intellect	
Low	High	Low	High	Low	High	Low	High	Low	High
−.83 Quiet	.85 Talkative	−.52 Fault-finding	.87 Sympathetic	−.58 Careless	.80 Organized	−.39 Stable*	.73 Tense	−.74 Commonplace	.76 Wide interests
−.80 Reserved	.83 Assertive	−.48 Cold	.85 Kind	−.53 Disorderly	.80 Thorough	−.35 Calm*	.72 Anxious	−.73 Narrow interests	.76 Imaginative
−.75 Shy	.82 Active	−.45 Unfriendly	.85 Appreciative	−.50 Frivolous	.78 Planful	−.21 Contented*	.72 Nervous	−.67 Simple	.72 Intelligent
−.71 Silent	.82 Energetic	−.45 Quarrelsome	.84 Affectionate	−.49 Irresponsible	.78 Efficient	.14 Unemotional*	.71 Moody	−.55 Shallow	.73 Original
−.67 Withdrawn	.82 Outgoing	−.45 Hard-hearted	.84 Soft-hearted	−.40 Slipshot	.73 Responsible		.71 Worrying	−.47 Unintelligent	.68 Insightful
−.66 Retiring	.80 Outspoken	−.38 Unkind	.82 Warm	−.39 Undependable	.72 Reliable		.68 Touchy		.64 Curious
	.79 Dominant	−.33 Cruel	.81 Generous	−.37 Forgetful	.70 Dependable		.64 Fearful		.59 Sophisticated
	.73 Forceful	−.31 Stern*	.78 Trusting		.68 Conscientious		.63 High-strung		.59 Artistic
	.73 Enthusiastic	−.28 Thankless	.77 Helpful		.66 Precise		.63 Self-pitying		.59 Clever
	.68 Show-off	−.24 Stingy*	.77 Forgiving		.66 Practical		.60 Temperamental		.58 Inventive
	.68 Sociable		.74 Pleasant		.65 Deliberate		.59 Unstable		.56 Sharp-witted
	.64 Spunky		.73 Good-natured		.46 Painstaking		.58 Self-punishing		.55 Ingenious
	.64 Adventurous		.73 Friendly		.26 Cautious*		.54 Despondent		.45 Witty*
	.62 Noisy		.72 Cooperative				.51 Emotional		.45 Resourceful*
	.58 Bossy		.67 Gentle						.37 Wise
			.66 Unselfish						.33 Logical*
			.56 Praising						.29 Civilized*
			.51 Sensitive						.22 Foresighted*
									.21 Polished*
									.20 Dignified*

Note. These 112 items were selected as initial prototypes for the Big Five because they were assigned to one factor by at least 90% of the judges. The factor loadings, shown for the hypothesized factor, were based on a sample of 140 males and 140 females, each of whom had been described by 10 psychologists serving as observers during an assessment weekend at the Institute of Personality Assessment and Research at the University of California at Berkeley (John, 1990).

*Potentially misclassified items (i.e., loading more highly on a factor different from the one hypothesized in the original prototype definition).

analyses suggest an interpretation closer to Intellectual Interests or even Openness than to the original interpretation as Culture (Norman, 1963). In Goldberg's (1990) factor analysis of Norman's 75 categories, Factor V was defined by Originality, Wisdom, Objectivity, Knowledge, Reflection, and Art, thus involving facets of Openness related to ideas, fantasy, and aesthetics (Costa & McCrae, 1992). When the 133 synonym clusters were factored, the two clusters labeled Intellectuality (intellectual, contemplative, meditative, philosophical, and introspective) and Creativity (creative, imaginative, inventive, ingenious, innovative) had the highest loadings, followed by Intelligence, Versatility, Wisdom, Perceptiveness, Art, Logic, Curiosity, and Nonconformity. The variables related to Cultural Sophistication (cultured, refined, worldly, cosmopolitan, urbane) did not load consistently on Factor V, and Dignity (mannerly, dignified, formal) loaded more highly on Conscientiousness than on Factor V. Nonconformity (nonconforming, unconventional, rebellious) loaded positively, and Conventionality (traditional, conventional, unprogressive) loaded negatively on Factor V in all four samples. These findings are inconsistent with the Culture interpretation and instead favor an Openness interpretation (McCrae, 1996). The finding that Unconventionality and Nonconformity load on Factor V is also consistent with the definition of this factor in Dutch and Italian (De Raad, Perugini, et al., 1998).

Indeed, Peabody and Goldberg (1989) concluded that the initial interpretation of Tupes and Christal's (1961) fifth factor as Culture was a historical accident. Peabody and Goldberg compared their representative variable selection with Cattell's and found that his selection underrepresented traits related to intellectual interests and overrepresented traits related to Culture. Even in Norman's (1963) studies, only one of the four variables included as a marker of Factor V was a measure of Cultural Sophistication: "polished, refined versus crude, boorish." The other three variables ("Artistically sensitive versus insensitive"; "Intellectual versus unreflective, narrow"; "Imaginative versus simple, direct") have more to do with creativity, cognitive complexity, and broad interests (i.e., Openness) than with being cultured, well-educated, and from an upper-class background. In 1963 as much as today, Factor V seems to encompass a broad range of intellectual, creative, and artistic inclinations, preferences, and skills found foremost in highly original and creative individuals (Barron, 1968; Helson, 1967; Gough, 1979; MacKinnon, 1965).

An alternative label for Factor V is Intellect. For example, Peabody and Goldberg (1989) included both *controlled* aspects of intelligence (perceptive, reflective, intelligent) and *expressive* aspects (imaginative, curious, broad-minded). The Intellect interpretation emphasizes thinking and reasoning but omits aspects of thought and experience that reflect personal orientations and attitudes, such as aesthetic and artistic interests, nonconformity, and progressive values. Indeed, the fifth factor is *not* a measure of intelligence, and it has only small positive correlations with measures of IQ and scholastic aptitude (e.g., Helson, 1985; John, Caspi, Robins, Moffitt, & Stouthamer-Loeber, 1994; McCrae & Costa, 1985c). Moving away from a narrow Intellect interpretation, Saucier (1994b) has suggested the label Imagination, which is somewhat closer to Openness and emphasizes that fantasy, ideas, and aesthetics, rather than intelligence, are most central to this factor.

In our view, Intellect is one part of a broader personality factor that McCrae and Costa (1985b, 1987) have described as Openness to Experience. Nonetheless, there is still some debate about the best interpretation of the fifth factor, and a special issue of the *European Journal of Personality* was devoted to this topic (see De Raad, 1994).

The Big Five Inventory (BFI): Measuring the Core Features of the Big Five with Short Phrases

To address the need for a short instrument measuring the prototypical components of the Big Five that are common across studies, John, Donahue, and Kentle (1991) constructed the Big Five Inventory (BFI; see also Benet-Martinez & John, 1998). The 44-item BFI was developed to represent the prototype definitions developed through expert ratings and subsequent factor analytic verification in observer personality ratings (see Table 4.2). The goal was to create a brief inventory that would allow efficient and flexible assessment of the five dimensions when there is no need for more differentiated measurement of individual facets. There is much to be said in favor of brevity; as Burisch (1984) observed, "Short scales not only save testing time, but also avoid subject boredom and fatigue . . . there are subjects . . . from whom you won't get any response if the test looks too long" (p. 219).

The BFI does not use single adjectives as items because such items are answered less consistently than when they are accompanied by definitions or elaborations (Goldberg & Kilkowski, 1985). Instead, the BFI uses short phrases based on the trait adjectives known to be prototypical markers of the Big Five (John, 1989, 1990). One or two prototypical trait adjectives served as the item core to which elaborative, clarifying, or contextual information was added. For example, the Openness adjective "original" became the BFI item "Is original, comes up with new ideas," and the Conscientiousness adjective "persevering" served as the basis for the item "Perseveres until the task is finished." Thus the BFI items (which are reprinted here in the Appendix) retain the advantages of adjectival items (brevity and simplicity) while avoiding some of their pitfalls (ambiguous or multiple meanings and salient desirability).

Although the BFI scales include only eight to ten items, they do not sacrifice either content coverage or good psychometric properties. For example, the nine-item Agreeableness scale includes items related to at least five of the six facets postulated by Costa and McCrae (1992)—namely, Trust (forgiving, trusting), Altruism (helpful and unselfish), Compliance (not quarrelsome), Modesty (not faultfinding with others), and Tender-Mindedness (considerate and kind). In U.S. and Canadian samples, the alpha reliabilities of the BFI scales typically range from .75 to .90 and average above .80; three-month test–retest reliabilities range from .80 to .90, with a mean of .85. Validity evidence includes substantial convergent and divergent relations with other Big Five instruments as well as with peer ratings.

MEASUREMENT: COMPARING THREE BIG FIVE INSTRUMENTS

So far, we have discussed Goldberg's (1992) TDA, Costa and McCrae's (1992) NEO questionnaires, and the BFI. In addition, a variety of other measures are available to assess the Big Five in English. Most of them were developed for specific research applications. Digman (e.g., 1963, 1989) constructed several different adjective sets to study teacher ratings of personality in children and adolescents. Big Five scales have also been constructed using items from existing instruments. For example, John and colleagues (1994) developed scales to measure the Big Five in adolescents using personality ratings on the California Child Q-sort obtained from their mothers. In their behavior genetic research, Loehlin, McCrae, Costa, and John (1998) used Big Five scales specifically constructed from the California Psychological Inventory (Gough, 1987) and the Adjective Check List (Gough & Heilbrun, 1983). Another broad-band personality inventory that provides scores for the Big Five is the Hogan Personality Inventory (Hogan, 1986). Extraversion is represented by the Sociability and Ambition scales, Agreeableness is represented by Likeability, Conscientiousness by Prudence (vs. impulsivity), Neuroticism by low scores on Adjustment, and Openness by Intellectance (see Table 4.5 later on in this chapter). The availability of so many different instruments to measure the Big Five makes clear that there is no single instrument that represents *the* gold standard.

Comparing the TDA, NEO-FFI, and BFI

In general, the NEO questionnaires represent the best-validated Big Five measures in the questionnaire tradition. Goldberg's (1992) 100-item TDA is the most commonly used measure consisting of single adjectives. Finally, the BFI has been used frequently in research settings in which subject time is at a premium and the short-phrase item format provides more context than Goldberg's single adjective items but less complexity than the sentence format used by the NEO questionnaires.

How well do these different Big Five measures converge? Moreover, are the five dimensions really independent? Critics have suggested that some of the Big Five dimensions are highly intercorrelated (Block, 1995; Eysenck, 1992). How high are these intercorrelations, and do they involve the same dimensions across instruments?

A number of studies have reported on the psychometric characteristics of each instrument, and a few studies have compared two instruments with each other (e.g., Benet-Martinez & John, 1998; Goldberg, 1992; McCrae & Costa, 1987). However, no published studies have compared all three. To provide such a comparison, we summarize findings from a large data set of self-reports on all three measures. The sample consisted of 462 undergraduates (61% female) at the University of California, Berkeley, who completed the TDA, the NEO-FFI, and the BFI. We analyzed the data in a multitrait–multimethod (MTMM) design in which the meth-

ods are three self-report instruments rather than different data sources (for a recent review, see John & Benet-Martinez, in press).

Although we expected the convergent validities across the three instruments to be substantial, we have already noted some subtle but important differences in the definitions of Extraversion and Openness. The NEO definition of Extraversion in terms of six facets was already in place before Costa and McCrae added domain scales for Agreeableness and Conscientiousness in 1985 and facet scales for these two factors in 1992. The Warmth facet scale, included in Extraversion (see Table 4.1), also correlates with their Agreeableness domain scale (Costa & McCrae, 1992). In contrast, Goldberg (1992) and John (1990) found that trait adjectives related to Warmth correlate more highly with Agreeableness than with Extraversion, suggesting that Warmth should be included on Agreeableness (see Table 4.2). The other potential difference involves the fifth factor. As described above, Goldberg (1992) interprets it as Intellect or Imagination (Saucier, 1992), thus emphasizing Openness to Ideas and to Fantasy over the other four facets. Similarly, the BFI Openness scale does not include items related to Costa and McCrae's (1992) Values and Actions facets. In college student samples, preliminary BFI items intended to measure liberal versus conservative values (for the Values facet), and behavioral flexibility (for the Action facet) failed to cohere with the other items on the BFI Openness scale (John et al., 1991).

Reliability of the Three Instruments

The coefficient alpha reliabilities are given in Table 4.3. Overall, the reliabilities were impressive for these relatively short scales. Not surprisingly, the longer TDA scales had the highest alphas (mean of .89), followed by the BFI (.83) and the NEO-FFI (.79). Across instruments, Extraversion, Conscientiousness, and Neuroticism were measured most reliably, whereas Agreeableness and Openness tended to be less reliable. The scale with the lowest reliability was the NEO-FFI Openness scale, replicating a finding in a different sample (Benet-Martinez & John, 1998). A number of NEO-FFI Openness items did not correlate well with the total scale in this student sample. These less reliable items included both of the items from the Action facet, as well as both of the Values items. It is possible that on liberal college campuses, items involving trying new and foreign foods (Action) and looking to religious authorities for decisions on moral issues (reverse scored on Values) do not discriminate as well as in Costa and McCrae's (1992) samples of older adults. In contrast, the three items from the Ideas facet (e.g., intellectual curiosity) and the three items from the Aesthetics facet (e.g., experiential effects of poetry or art) had the strongest item-total correlations. Finally, in contrast to the heavy representation of imagination items on the TDA, only one item related to imagination (from the Fantasy facet) was included on the NEO-FFI Openness scale.[3]

Convergent and Discriminant Validity across the Three Instruments

As a first test of cross-instrument convergence, we examined the full 15 × 15 MTMM correlation matrix formed by the five factors crossed with the three instruments.[4] In general, the cross-instrument validity correlations, computed between pairs of instruments and shown in Table 4.3, were substantial. Across all five factors, the mean of the convergent validity correlations across instruments was .75. As shown in Table 4.3, the BFI and TDA showed the strongest convergence (mean r = .81), followed by the BFI and NEO-FFI (mean r = .73), and finally the TDA and NEO-FFI (mean r = .68).[5]

To determine the extent to which the validity correlations simply reflect the imperfect reliability of the scales rather than substantive differences among the instruments, we corrected for attenuation using alpha. As shown in Table 4.3, the corrected validity correlations averaged .91. However, this excellent overall result masks some important differences. Across instruments, the first three of the Big Five (Extraversion, Agreeableness, and Conscientiousness) showed mean validities clearly exceeding .90, suggesting virtual equivalence among the instruments. Neuroticism (.88) and Openness (.83) were lower. Focusing on the pairwise comparisons between instruments, the BFI and the TDA shared virtually all of their reliable variance (corrected mean r = .95). Convergence between the BFI and the NEO-FFI was also substantial (mean = .93); however, the correlations for Extraversion and for Openness did not reach .90, suggesting that the conceptualizations of these factors are not fully equivalent across these two instruments. A similar pattern was observed for the TDA and the NEO-FFI but the convergent correlations were generally lower (mean = .83) and fell below

TABLE 4.3. Reliability and Convergent Validity Coefficients for the TDA, NEO-FFI, and BFI

	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness	Mean
			Reliabilities			
TDA	.92	.90	.90	.85	.88	.89
BFI	.88	.79	.82	.84	.81	.83
NEO	.78	.78	.83	.85	.70	.79
Mean	.87	.83	.85	.85	.81	**.84**
			Uncorrected pairwise convergent validities			
BFI-TDA	.90	.78	.81	.76	.75	.81
BFI-NEO	.69	.76	.79	.76	.64	.73
TDA-NEO	.67	.68	.77	.70	.56	.68
Mean	.78	.74	.79	.74	.66	.75
			Corrected pairwise convergent validities			
BFI-TDA	.99	.93	.94	.90	.89	.95
BFI-NEO	.83	.97	.96	.90	.85	.92
TDA-NEO	.79	.81	.89	.82	.71	.81
Mean	.93	.92	.94	.88	.83	**.91**
			Standardized validity coefficients from CFA (Model 6)			
BFI	.94	.92	.92	.90	.92	.92
TDA	.95	.85	.87	.83	.79	.87
NEO	.68[a]	.83	.86	.84	.70	.79
Mean	.90	.87	.89	.86	.83	**.87**

Note. $N = 462$. BFI, Big Five Inventory; TDA, Trait Descriptive Adjectives, NEO, NEO Five Factor Inventory. Grand means are shown in **bold**. All means are based on Fisher *r*-to-*Z* transformations.

[a]The NEO Extraversion scale had a cross-loading on Agreeableness in Model 6 (see Figure 4.1).

.80 for Extraversion and Openness. In short, the NEO-FFI showed greater convergence with the BFI than with the TDA, but it defined Extraversion and Openness somewhat differently than those two instruments.

Overall, discriminant correlations were low; absolute values averaged .21 for the TDA, .17 for the NEO-FFI, and .20 for the BFI. Moreover, none of the discriminant correlations reached .40 on any of the instruments, and the largest correlations were .39 for the TDA, .38 for the NEO-FFI, and .33 for the BFI. Averaged across instruments, only four of the 10 discriminant correlations exceeded .20: the mean correlation was .28 between Agreeableness and Conscientiousness, −.28 between Agreeableness and Neuroticism, −.27 between Extraversion and Neuroticism, and .24 between Extraversion and Conscientiousness. Thus, there was little support for Eysenck's (1992) contention that Agreeableness and Conscientiousness are highly correlated "primary" traits that combine into a broader dimension contrasting Eysenck's Psychoticism with what might be called "good character." Together the findings show that the Big Five are fairly independent dimensions that can be measured with convergent and discriminant validity.

Confirmatory Factor Analysis (CFA) of the Multitrait–Multimethod Matrix

As a more formal test of convergent and discriminant validity, we used a series of nested CFA models to estimate latent factors representing the Big Five, their intercorrelations, and method factors representing the unique characteristics of each instrument. The most basic model (see Model 1 in Table 4.4) specified five uncorrelated latent trait factors and no method factors. This model showed marginal fit. Allowing intercorrelations among the Big Five factors significantly improved model fit (Model 2) suggesting that some of the Big Five intercorrelations were consistent across all three instruments.

TABLE 4.4. Nested Confirmatory Factor Analyses of the Multitrait–Multimethod Matrix Formed by the Three Big Five Instruments

Model	χ^2	*df*	CFI	$\Delta\chi^2$
1. Uncorrelated Big Five, no method factors	783	90	.863	—
2. Correlated Big Five, no method factors	583	80	.900	200*
3. Correlated Big Five, no method factors, cross-loading of NEO-Extraversion on Agreeableness	496	79	.917	87*
4. Model 3 plus NEO method factor	484	74	.919	12*
5. Model 4 plus TDA method factor	323	69	.950	160*
6. Model 5 plus BFI method factor	296	64	.954	28*
6a. Model 6 plus cross-loading of TDA—Intellect on Conscientiousness	274	63	.958	22*

Note. $N = 462$. CFI, comparative fit index (Bentler, 1990); $\Delta\chi^2$, increase in overall fit; BFI, Big Five Inventory; NEO, NEO Five Factor Inventory; TDA, trait descriptive adjectives.

* $p < .05$.

As we noted earlier, the NEO-FFI includes items related to warmth in Extraversion, whereas the BFI and TDA include them in Agreeableness. We examined this hypothesis directly by modifying our model and testing the improvement in model fit. By adding a cross-loading of the NEO Extraversion scale on the latent Agreeableness factor, we achieved a significant improvement in fit (see Model 3). As shown in Figure 4.1, the NEO-FFI Extraversion scale still loaded substantially on the Extraversion factor but also had a secondary loading on Agreeableness.[6]

The next three models were increasingly complex, adding method factors specific to each instrument. In Model 4 we added a method factor for the NEO-FFI, producing a small improvement in fit; as shown in Figure 4.1, this method factor primarily represented instrument-specific variance related to Openness. Model 5 added a TDA method factor, yielding a sizeable improvement; this method factor represented a positive correlation between Agreeableness and Intellect observed on the TDA but not on the BFI and the NEO-FFI. Model 6 added a BFI method factor, modeling a moderate negative correlation between Neuroticism and Openness on the BFI that was not observed on the TDA and the NEO. In short, these method factors capture specific differences in the ways the Big Five dimensions are conceptualized on each of the instruments. Figure 4.1 shows the parameter estimates for Model 6, which accounts for trait variance, method variance, and the expected cross-loading of NEO Extraversion on the Agreeableness factor.

We also explored how we could improve fit further. When we examined the residual matrix to see what relationships were still unexplained by our model, we found that the largest unexplained covariances were between the TDA fifth factor scale (Intellect) and the three Conscientiousness scales. McCrae and Costa (1985c, 1987) had previously noted that Goldberg's conceptualization of Factor V as Intellect is related to Conscientiousness, whereas the Openness conceptualization on the NEO-FFI and BFI is not. When we respecified our model with a cross-loading of the TDA Intellect scale on the Conscientiousness factor (Model 6a), we did observe an improvement in fit, but it was very small and the estimated cross-loading was only .15. In our view, such small gains in fit do not justify the added complexity required by the more detailed model, leading us to prefer the more parsimonious model represented in Figure 4.1.

Table 4.3 summarizes the standardized validity coefficients from the CFA. They average .92 for the BFI, .87 for the TDA, and .79 for the NEO-FFI, suggesting that the canonical representation achieved by the CFA is captured most closely by the BFI, which was developed to capture the core characteristics of the Big Five. The parameter estimates for Model 6 (see Figure 4.1) suggest three major conclusions that are consistent with the preceding analyses. First, all fifteen scales had substantial loadings on the five latent factors, with an average loading of .87, suggesting that all three measures generally tap the same five dimensions. Second, the substantial size of these loadings did not leave

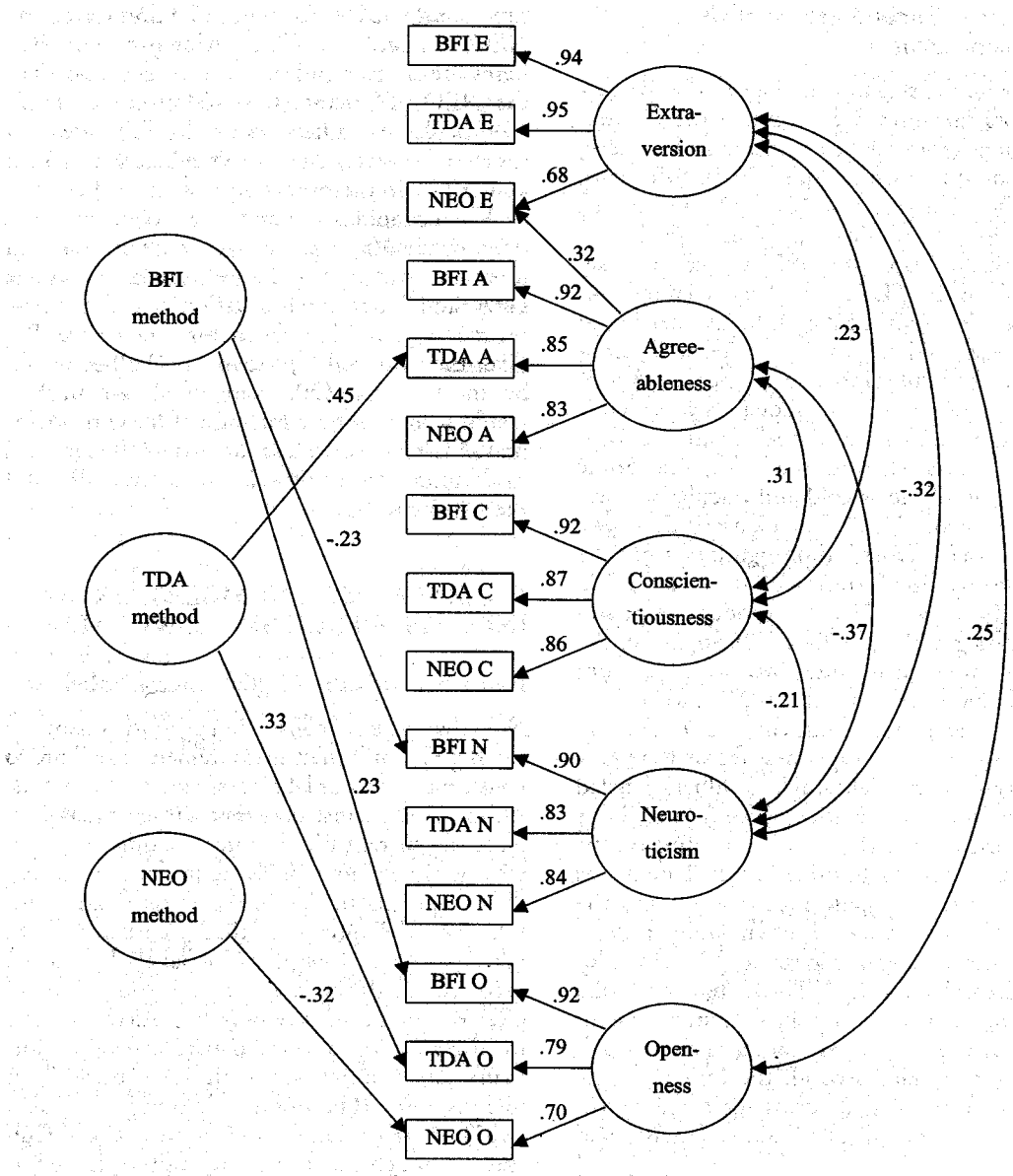

FIGURE 4.1. Standardized parameter estimates for the final multitrait–multimethod model (Model 6 in Table 4.3). Method effects and trait intercorrelations less than .20 and error terms are not shown.

much systematic variance for general instrument factors; instead, the three latent method factors we did uncover related to specific scale intercorrelations that were unique to each instrument. Nonetheless, in all cases the loadings on these method factors were considerably smaller than the substantive trait loadings, suggesting that the measures are more similar than different. The third conclusion involves the size of the intercorrelations among the latent Big Five dimensions, which remained low even when disattenuated for unreliability by CFA; none of them reached .40. Overall, then, the CFA results show that five latent, modestly correlated personality factors capture the major sources of variance in our MTMM design, and three smaller method factors represent trait-specific variance for each instrument.

A Joint Item Factor Analysis of the Three Instruments

To elaborate the shared meanings of the five factors across measures, we examined the highest-loading items for each factor in a joint item-level factor analysis, which included all 44 BFI items, 60 NEO-FFI items, and 100 TDA items. For Extraversion, the top-loading items were "Is outgoing, sociable" from the BFI, "Quiet" (reverse-scored) from the TDA, and "I really enjoy talking to people" from the NEO. Items referring to assertiveness, activity level, and positive emotions also had substantial loadings. For Agreeableness, item examples included "Is considerate and kind to almost everyone" from the BFI, "Unkind" (reversed) from the TDA, and "Some people think of me as cold and calculating" (reversed) from the NEO. For Conscientiousness, key items were "Does a thorough job" from the BFI, "Disorganized" (reversed) from the TDA, and the NEO item "I am a productive person who always gets the job done." Exemplars of the Neuroticism factor include "Worries a lot" from the BFI, "Nervous" from the TDA, and "I often feel tense and jittery" from the NEO. The top loadings on the joint Openness factor were particularly instructive: although Goldberg labeled his scale Intellect (or Imagination), the TDA item "Creative" had the strongest loading on the joint factor. The highest-loading BFI item was "Values artistic, aesthetic experiences," and the best NEO items were "I often enjoy playing with theories or abstract ideas" and "I have a lot of intellectual curiosity." These item examples for Openness make two points. First, the factor clearly involves Openness rather than intellectual ability or skill. Second, the aspects of the Openness factor shared across the three instruments involve openness to ideas, fantasy, and aesthetics.

Big Five Measurement: Conclusions and Limitations

One of the limitations of the findings presented here is that we did not examine external (or predictive) validity. Both the NEO questionnaires and the BFI have been shown to predict peer ratings; such evidence still needs to be obtained for the TDA scales. Future research needs to compare the validity of all three instruments using peer ratings and other external criteria. One of the advantages of the BFI is its efficiency, taking only about 5 minutes of administration time, compared with about 15 minutes for the NEO-FFI and the TDA. Moreover, the BFI items are shorter and easier to understand than the NEO-FFI items (Benet-Martinez & John, 1998). The 100 adjectives on the TDA are even shorter; however, single-trait adjectives can be ambiguous in their meanings (see note 3).

When should researchers use each of these instruments? When participant time is not at a premium, participants are well educated and test savvy, and the research question calls for the assessment of multiple facets for each of the Big Five, then the full 240-item NEO PI-R would be most useful. Otherwise, the 44-item BFI would seem to offer a measure of the core attributes of the Big Five that is at least as efficient and easily understood as the 60-item NEO-FFI and the 100-item TDA.

FACTOR NAMES, NUMBERS, OR INITIALS: WHICH SHALL WE USE?

Problems with the English Factor Labels

Now that we have considered both the history of the Big Five and their measurement, it is time to revisit the names or labels assigned to the factors. Although the constructs that will eventually replace the current Big Five may be different from what we know now, labels are important because they imply particular interpretations and thus influence the directions that theorizing might take. Norman's (1963) factor labels have been used frequently in later research, but Norman offered little in the way of a theoretical rationale for the selection of these particular labels. Norman's labels differ vastly in their breadth or inclusiveness (Hampson, Goldberg & John, 1987); in particular, Conscientiousness and Culture are much too narrow to capture the enormous breadth of these two dimensions. Moreover, as noted above, researchers quickly abandoned Culture as a label for Factor V, in favor of Intellect or Imagination (Saucier & Goldberg, 1996a) or Openness to Experience (McCrae & Costa, 1985b). Neither label is truly satisfactory, however, because Intellect is too narrow and Openness, while broad enough, is somewhat vague.

Agreeableness is another problematic label. For one, it refers to the behavioral tendency to *agree* with others, thus incorrectly implying submissiveness, which is more closely related to the introverted pole of Factor I. Agreeableness is also

too detached, too neutral a label for a factor supposed to capture intensely affective characteristics, such as love, compassion, and sympathy. Freud viewed love and work as central; following this lead, we could call Factor II simply Love (Peabody & Goldberg, 1989).

However, Work is too narrow a label for Factor III. Even Conscientiousness is too narrow because it omits a central component that Peabody and Goldberg (1989) called "favorable impulse control." Thus, Responsibility or even Degree of Socialization (see Gough, 1987) might be labels more appropriate for Factor III than is Conscientiousness.

More could be said about the many shortcomings of the traditional labels (see also Block, 1995), but better labels are hard to come by. The unsurpassed advantage of the traditional labels is that they are commonly known and used, thus preventing Babel from taking over the literature on the Big Five. Moreover, before any new names are devised, the definition of the factors in terms of facets or components must be elaborated and sharpened. At this point, it seems premature to settle the scope and theoretical interpretation of the factors by devising new names.

Preliminary Definitions

Because the traditional labels are so easily misunderstood, short definitions of the five dimensions may be useful here (cf., Costa & McCrae, 1992; John, 1990; Tellegen, 1985). Briefly, Extraversion implies an *energetic approach* to the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality. Agreeableness contrasts a *prosocial and communal orientation* toward others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty. Conscientiousness describes *socially prescribed impulse control* that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks. Neuroticism contrasts emotional stability and even-temperedness with *negative emotionality*, such as feeling anxious, nervous, sad, and tense. Finally, Openness to Experience (versus closed-mindedness) describes the breadth, depth, originality, and complexity of an individual's *mental and experiential life*.

The numbering convention from I to V, favored by Saucier and Goldberg (1996b) and Hofstee and colleagues (1997), is useful because it reflects the relative size of the factors in lexical studies. Factor I and II, which primarily summarize traits of interpersonal nature, tend to account for the largest percentage of variance in personality ratings, followed by Factor III, whereas the last two factors are by far the smallest in lexical studies (De Raad, Perugini, Hrebickova, & Szarota, 1998). However, the Roman numerals are hard to remember, and the order of the factors is not invariant across studies. Thus, we favor the mnemonic convention suggested by the initials given below. They evoke multiple associations that represent more fully than a single word the broad range of meaning captured by each of the factors:

- **E** Extraversion, Energy, Enthusiasm (I)
- **A** Agreeableness, Altruism, Affection (II)
- **C** Conscientiousness, Control, Constraint (III)
- **N** Neuroticism, Negative Affectivity, Nervousness (IV)
- **O** Openness, Originality, Open-Mindedness (V)

The reader intrigued by anagrams may have noticed that these letters form the OCEAN of personality dimensions.

CONVERGENCE BETWEEN THE BIG FIVE AND OTHER STRUCTURAL MODELS

McCrae and Costa's (1985a, 1985b, 1985c; 1987) findings, like evidence for cross-instrument convergence presented above, show that the factor-analytic results from the lexical tradition converge surprisingly well with those from the questionnaire tradition. This convergence has led to a dramatic change in the acceptance of the five factors in the field. With regard to their empirical status, the findings accumulated since the mid-1980s show that the five factors replicate across different types of subjects, raters, and data sources, in both dictionary-based and questionnaire-based studies. Indeed, even skeptical reviewers were led to conclude that "agreement among these descriptive studies with respect to *what* are the appropriate dimensions is impressive" (Revelle, 1987, p. 437; see also Briggs, 1989; McAdams, 1992; Pervin, 1994). The finding that it doesn't matter whether Conscien-

tiousness is measured with trait adjectives, short phrases, or questionnaire items suggests that the Big Five dimensions have the same conceptual status as other personality constructs. For example, Loehlin and colleagues (1998) found that all five factors show substantial and about equal heritabilities, regardless of whether they are measured with questionnaires or with adjective scales derived from the lexical approach.

One of the apparent strengths of the Big Five taxonomy is that it can capture, at a broad level of abstraction, the commonalities among most of the existing systems of personality traits, thus providing an integrative descriptive model for research. Table 4.5 summarizes the personality dimensions proposed by a broad range of personality theorists and researchers. These dimensions, although by no means a complete tabulation, emphasize the diversity of current conceptions of personality. However, they also point to some important convergences. First, almost every one of the theorists includes a dimension akin to Extraversion. Although the labels and exact definitions vary, nobody seems to doubt the fundamental importance of this dimension (Guilford, 1974, 1975). The second almost universally accepted personality dimension is Emotional Stability, as contrasted with Neuroticism, Negative Emotionality, and Proneness to Anxiety (Tellegen, 1982, 1985). Interestingly, however, not all the researchers listed in Table 4.5 include a separate measure for this dimension. This is particularly true of the interpersonal researchers, such as Wiggins (1979) and Bales (1970), as well as the questionnaires aimed primarily at the assessment of basically healthy, well-functioning adults, such as Gough's (1987) CPI, the Myers–Briggs Type Indicator (Myers & McCaulley, 1985), and even Jackson's (1984) PRF. In contrast, all of the temperament-based models include Neuroticism. There is somewhat less agreement on the third dimension, which appears in various guises, such as Control, Constraint, Super-Ego Strength, and Work Orientation as contrasted with Impulsivity, Psychoticism, and Play Orientation. The theme underlying most of these concepts involves the control, or moderation, of impulses in a normatively and socially appropriate way (cf. Block & Block, 1980). However, Table 4.5 also points to the importance of Agreeableness and Openness, which are neglected by temperament-oriented theorists such as Buss and Plomin (1975) and Eysenck (1985). In a comprehensive taxonomy, even at the broadest level, we need a "place" for an interpersonal dimension related to Communion, Feeling Orientation, Altruism, Nurturance, Love Styles, and Social Closeness, as contrasted with Hostility, Anger Proneness, and Narcissism. The existence of these questionnaire scales, and the cross-cultural work on the interpersonal origin and consequences of personality, stress the need for a broad domain akin to Agreeableness, Warmth, or Love.

Similar arguments apply to the fifth and last factor included in the Big Five. For one, there are the concepts of Creativity, Originality, and Cognitive Complexity, which are measured by numerous questionnaire scales (Barron, 1968; Helson, 1967, 1985; Gough, 1979). Although these concepts are cognitive, or, more appropriately, *mental* in nature, they are clearly different from IQ. Second, limited-domain scales measuring concepts such as Absorption, Fantasy Proneness, Need for Cognition, Private Self-Consciousness, Independence, and Autonomy would be difficult to subsume under Extraversion, Neuroticism, or Conscientiousness. Indeed, the fifth factor is necessary because individual differences in intellectual and creative functioning underlie artistic interests and performances, inventions and innovation, and even humor. Individual differences in these domains of human behavior and experience cannot be, and fortunately have not been, neglected by personality psychologists.

Finally, the matches between the Big Five and other constructs sketched out in Table 4.5 should be considered with a healthy dose of skepticism. Some of these correspondences are indeed based on solid research findings. Others, however, are conceptually derived and seem plausible, but await empirical confirmation. All of these matches reflect broad similarities, ignoring some important, implicative, and useful differences among the concepts proposed by different investigators. Nonetheless, at this stage in the field, we are more impressed by the newly apparent similarities than by the continuing differences among the various models. Indeed, the Big Five are useful primarily because of their integrative and heuristic value, a value that becomes apparent in Table 4.5. The availability of a taxonomy, even one that is as broad and incomplete as the Big Five, permits the comparison and potential integration of dimensions that, by their names alone, would seem entirely disparate.

TABLE 4.5. The Big Five and Dimensions of Similar Breadth in Questionnaires and in Models of Personality and Interpersonal Behavior

Theorist	Extraversion I	Agreeableness II	Conscientiousness III	Neuroticism IV	Openness/Intellect V
Bales	Dominant-Initiative	Social-Emotional Orientation	Task Orientation$_R$[a]		—
Block	Undercontrol	Overcontrol		Resiliency$_R$[b]	
Buss & Plomin	Activity	—	Impulsivity$_R$	Emotionality	—
Cattell	Exvia (vs. Invia)	Pathemia (vs. Cortertia)	Superego Strength	Adjustment$_R$ (vs. Anxiety)	Independence
Comrey Scales (Noller et al.)	Extraversion and Activity	Femininity (vs. Masculinity)	Orderliness and Social Conformity	Emotional Stability$_R$	Rebelliousness
Eysenck	Extraversion	Psychoticism$_R$[c]		Neuroticism	—
Gough					
CPI Vectors	Externality	—	Norm-Favoring	Self-Realization$_R$[d]	
CPI Scales	Sociability	Femininity	Norm-Favoring	Well-being	Achievement via Independence
Guilford	Social Activity	Paranoid-Disposition$_R$	Thinking Introversion	Emotional Stability$_R$	—
Hogan	Sociability	Likeability	Prudence (vs. Impulsivity)	Adjustment$_R$	Intellectance
Jackson	Outgoing, Social Leadership	Self-Protective Orientation$_R$	Work Orientation	Dependence	Aesthetic-Intellectual
MMPI Personality Disorder Scales	Histrionic	Paranoid$_R$	Compulsive	Borderline	Schizotypal
Myers–Briggs	Extraversion (vs. Introversion)	Feeling (vs. Thinking)	Judging (vs. Perception)	—	Intuition (vs. Sensing)
Tellegen	Positive Emotionality		Constraint	Negative Emotionality	Absorption
	Agentive	Communal			
Wiggins[e]	Dominance	Nurturance	(Conscientiousness)	(Neuroticism)	(Openness)

Note. Based on John (1990) and McCrae and John (1992). Subscript "R" indicates that the dimension was reverse-scored in the direction *opposite* to that of the Big Five label listed above.

[a]This dimension contrasts a work-directed, emotionally neutral orientation with an erratic, emotionally expressive orientation (Bales & Cohen, 1979), and thus seems to combine elements of both Conscientiousness and Neuroticism.

[b]Resiliency seems to subsume aspects of both Openness and low Neuroticism, because an ego-resilient individual is considered both intellectually resourceful and effective in controlling anxiety (Block & Block, 1980). However, Robins, John, and Caspi (1994) found that in adolescents, ego-resiliency is related to all of the Big Five dimensions in the well-adjusted direction. Ego control was related to Extraversion, Conscientiousness, and Agreeableness, with undercontrol similar to Extraversion and overcontrol similar to Conscientiousness and Agreeableness.

[c]High scores on the EPQ Psychoticism scale are associated with low scores on both Agreeableness and Conscientiousness (Goldberg & Rosolack, 1994; McCrae & Costa, 1985a).

[d]The third vector scale on the CPI (Gough, 1987) measures levels of psychological integration and realization, and should reflect aspects of both low Neuroticism (e.g., Well-being) and high Openness (e.g., Achievement via Independence).

[e]Wiggins (1979) originally focused on Dominance and Nurturance, which define the interpersonal circumplex. Trapnell and Wiggins (1990) added adjective scales for Conscientiousness, Neuroticism, and Openness (see also Wiggins, 1995).

CRITICAL ISSUES AND THEORETICAL PERSPECTIVES

The Big Five provides a descriptive taxonomy that organizes the myriad natural-language and scientific trait concepts into a single classificatory framework. However, like any scientific model, it has limitations. Several critics have argued that the Big Five does not provide a complete *theory* of personality (e.g., Block, 1995; Eysenck, 1997; McAdams, 1992; Pervin, 1994). We agree. The Big Five taxonomy was never intended as a comprehensive personality theory; it was developed to account for the structural relations among personality traits (Goldberg, 1993). Thus, like most structural models it provides an account of personality that is primarily descriptive rather than explanatory, emphasizes regularities in behavior rather than inferred dynamic and developmental processes, and focuses on variables rather than on individuals or types of individuals (cf. John & Robins, 1993, 1998). Nonetheless, the Big Five trait taxonomy provides a conceptual foundation that helps examine these theoretical issues. In this section, we begin with the hierarchical structure defined by the Big Five, and then review whether the Big Five dimensions predict important life outcomes, how they develop, how they combine into personality types, and how different researchers view their conceptual status.

Hierarchy, Levels of Abstraction, and the Big Five

A frequent objection to the Big Five is that five dimensions cannot possibly capture all of the variation in human personality (Block, 1995; Briggs, 1989; McAdams, 1992; Mershon & Gorsuch, 1988), and that they are much too broad. However, the objection that five dimensions are too few overlooks the fact that personality can be conceptualized at different levels of abstraction or breadth. Indeed, many trait domains are hierarchically structured (Hampson, John, & Goldberg, 1986).

The advantage of categories as broad as the Big Five is their enormous bandwidth. Their disadvantage, of course, is their low fidelity. In any hierarchical representation, one always loses information as one moves up the hierarchical levels. For example, categorizing something as a "guppy" is more informative than categorizing it as a "fish," which in turn is more informative than categorizing it as an "animal." Or, in psychometric terms, one necessarily loses item information as one aggregates items into scales, and one loses scale information as one aggregates scales into factors (John, Hampson, & Goldberg, 1991).

The Big Five dimensions represent a rather broad level in the hierarchy of personality descriptors. In that sense, they are to personality what the categories "plant" and "animal" are to the world of biological objects—extremely useful for some initial rough distinctions but of less value for predicting specific behaviors of a particular object. The hierarchical level a researcher selects depends on the descriptive and predictive tasks to be addressed (Hampson et al., 1986). In principle, the number of specific distinctions one can make in the description of an individual is infinite, limited only by one's objectives.

Norman, Goldberg, McCrae and Costa, and Hogan all recognized that there was a need in personality, just as in biology, "to have a system in which different levels of generality or inclusion are recognized" (Simpson, 1961, p. 12). A complete trait taxonomy must include middle-level categories, such as Assertiveness, Orderliness, and Creativity, and even narrower descriptors, such as talkative, punctual, and musical (John et al., 1991). Therefore Norman and, more extensively, Goldberg (1982, 1990) have developed between 40 and 75 middle-level categories subordinate to the Big Five dimensions (for a review, see John et al., 1988). However, as Briggs (1989) noted, Norman's and Goldberg's middle-level categories have not been investigated systematically nor have they been included in an assessment instrument. At this point, Costa and McCrae's (1992) 30 facets represent the most elaborated and empirically validated model. Hofstee and colleagues' (1992) circumplex-based approach, which defines facets as pairwise combinations of two factors, is another promising direction to pursue. However, the two approaches differ notably in the facets they propose, indicating the need for further conceptual and empirical work to achieve a consensual specification of the Big Five factors at this lower level of abstraction.

Predicting Important Life Outcomes

External validity and predictive utility are topics that in the past have received conspicuously little attention from researchers working in the Big Five tradition. However, one of the criteria for the usefulness of a structural model is its success

in predicting important outcomes in people's lives. Eysenck (1991) argued that "little is known about the social relevance and importance of Openness, Agreeableness, and Conscientiousness. . . . What is lacking is a series of large-scale studies which would flesh out such possibilities" (p. 785). According to Eysenck (1991), the validity of the Big Five should be examined against socially relevant criteria such as criminality, mental illness, academic aptitude and achievement, and work performance.

A large study of adolescents has addressed this challenge, examining three of Eysenck's criteria: juvenile delinquency, childhood psychopathology, and academic performance (see John et al., 1994; Robins et al., 1994). The findings suggest that the Big Five can help us understand theoretically, socially, and developmentally significant life outcomes. For example, low Agreeableness and low Conscientiousness predict juvenile delinquency. In terms of psychopathology, Neuroticism and low Conscientiousness predict internalizing disorders. Conscientiousness and Openness predict school performance. These findings suggest that the Big Five dimensions can be used as indicators of risk for subsequent maladjustment. Huey and Weisz's (1997) findings suggest that these links between personality and life outcomes hold up in a clinical sample as well. Researchers may eventually use Big Five profiles to identify children at risk and ultimately design appropriate interventions, such as teaching children low in Conscientiousness relevant behaviors and skills (e.g., strategies for delaying gratification).

The literature on adults also provides evidence for the external validity of the Big Five. For example, in studies of job performance (for reviews see Barrick & Mount, 1991; Mount, Barrick, & Stewart, 1998), the Big Five have been found to relate to important outcomes in the workplace. Conscientiousness has emerged as the only general predictor of job performance, although other dimensions relate to more specific aspects of job performance. For example, Agreeableness and Neuroticism predict performance in jobs in which employees work in groups, whereas Extraversion predicts success in sales and management positions. These trait-by-job interactions help researchers develop a more fine-grained understanding of how different traits are instrumental to performance in various job environments.

The availability of the Big Five taxonomy has also renewed interest in the links between personality and adult psychopathology (e.g., Wiggins & Pincus, 1989); findings from this burgeoning literature have been reviewed in Costa and Widiger (1994). The Big Five has also helped bring order to the many, often confusing, findings linking personality traits to physical health (see Adams, Cartwright, Ostrove, & Stewart, 1998; Friedman, Hawley, & Tucker, 1994; Friedman, Tucker, Schwartz, & Tomlinson-Keasey, 1995); the accumulated evidence now suggests that the regular and well-structured lives led by individuals high in Conscientiousness are conducive to better health outcomes and longevity, whereas antagonistic hostility (i.e., low Agreeableness) and negative affect (i.e., high Neuroticism) appear to be risk factors.

The emerging nomological network for each of the Big Five now includes an ever-broadening range of life outcome variables, such as leadership (Extraversion), helping others and donating to charity (Agreeableness), school and college grades (Conscientiousness), vulnerability to depression (Neuroticism), creative performance (Openness), and so on. These findings have been summarized in several recent reviews (Graziano & Eisenberg, 1997; Hogan & Ones, 1997; McCrae, 1996; Watson & Clark, 1997).

In interpreting these findings, it is important to realize that although personality traits are stable, people can change their patterns of behavior, thought, and feeling as a result of therapy and intervention programs (Heatherton & Weinberger, 1994). Thus, the links between the Big Five and important life outcomes point to behavioral domains that people can target for personal development and change; for example, people can improve how conscientiously they adhere to a diet, exercise regimen, or medical treatment plan (Friedman et al., 1994).

The Big Five and Personality Development

Historically, personality psychology has concerned itself with a range of developmental issues that are relevant to the Big Five—the antecedents of adult personality traits, how traits develop, the timelines for the emergence and peak expression of traits, their stability or change throughout the life span, and the effects of traits on other aspects of personal development. Some critics have suggested that Big Five researchers have not paid enough attention to issues of personality development in childhood and adolescence (Pervin, 1994). This criticism has some

merit: Although the Big Five taxonomy has influenced research on adult development and aging (Field & Millsap, 1991; Helson & Stewart, 1994; McCrae & Costa, 1990), there has been little research on personality structure in childhood. Developmental and temperament psychologists have studied a number of important traits (e.g., sociability, fearful distress, shyness, impulsivity) but they tend to study one trait at a time, in isolation from the others, and the available research has not been integrated in a coherent taxonomic framework. Until this work is done, however, research on personality development across the life span is likely to remain fragmented (Halverson, Kohnstamm, & Martin, 1994).

The adult personality taxonomy defined by the Big Five can offer some promising leads. In our view, the Big Five should be examined in developmental research for two reasons (John et al., 1994). Theoretically, it may be necessary to examine the developmental origins of the Big Five: Given that the Big Five emerge as basic dimensions of personality in adulthood, researchers need to explain how they develop. Practically, the Big Five taxonomy has proven useful as a framework for organizing findings on adult personality in areas as diverse as behavioral genetics and industrial psychology. Thus, extension of the Big Five into childhood and adolescence would facilitate comparisons across developmental periods.

Work on these issues has now begun, and researchers are drawing on existing models of infant and child temperament (see Clark & Watson, Chapter 16, this volume) to make connections to the Big Five dimensions in adulthood. A book edited by Halverson and colleagues (1994) summarizes these recent efforts. Some research suggests that the Big Five may provide a good approximation of personality structure in childhood and adolescence (Digman, 1989; Graziano & Ward, 1992). Extending Digman's (1989) earlier work on Hawaiian children, Digman and Shmelyov (1996) examined both temperament dimensions and personality dimensions in a sample of Russian children. Based on analyses of teachers' ratings, they concluded that the Big Five taxonomy offers a useful model for describing the structure of temperament. Studies using free-response techniques found that the Big Five can account for a substantial portion of children's descriptions of their own and others' personalities (Donahue, 1994), as well as teachers' and parents' descriptions of children's personality (Kohnstamm, Halverson, Mervielde, & Havill, 1998).

Two large-scale studies suggest that the picture may be more complicated. John et al. (1994) tested whether the adult Big Five structure would replicate in a large and ethnically diverse sample of adolescent boys. This research used the California Child Q-set (CCQ; Block & Block, 1969), a comprehensive item pool for the description of children and adolescents that was not derived from the adult Big Five and does not represent any particular theoretical orientation. Factor analyses identified five dimensions that corresponded closely with a priori scales representing the adult Big Five. However, two additional dimensions emerged in this study: "Irritability" was defined by items that involve negative affect expressed in age-inappropriate behaviors, such as whining, crying, tantrums, and being overly sensitive to teasing. "Activity" was defined by items involving physical activity, energy, and high tempo, such as running, playing, and moving and reacting quickly. In several Dutch samples of boys and girls aged 3 to 16 years, van Lieshout and Haselager (1994) also found the Big Five plus two factors similar to Irritability and Activity, thus supporting the generalizability of these dimensions across cultures and the two sexes. These replicated findings suggest that the structure of personality traits may be more differentiated in childhood than in adulthood. Specifically, the two additional dimensions may originate in temperamental features of childhood personality (i.e., irritable distress and activity level) that become integrated into adult personality structure over the course of adolescence (John et al., 1994).

These studies illustrate how the Big Five can help stimulate research that connects and integrates findings across long-separate research traditions. These studies also provide some initial insights about the way personality structure may develop toward its adult form. Yet, a great deal of work still lies ahead. Change in personality structure should be studied with reference to maturational changes, social-contextual transitions, and age-specific life tasks. Longitudinal research can help map changes in the dimensional structure of personality and discover how temperamental characteristics observed in infancy and early childhood manifest themselves during adolescence and adulthood. Finally, studies need to examine the antecedents of the Big Five and their relations to other aspects of personality functioning in childhood and adolescence. In

this way, the Big Five can help connect research on adult personality with the vast field of social development (Caspi, 1997).

Personality Types and Dynamics

The emergence of the Big Five has also rekindled interest in personality types. Note that the Big Five dimensions provide a model of personality structure that represents the covariation among personality traits *across* individuals. However, "personality structure" can also refer to the organization of traits *within* the individual (Allport, 1958). Person-centered research focuses on the particular configuration, patterning, and dynamic organization of the individual's total set of characteristics (cf. York & John, 1992; see also Magnusson, Chapter 8, this volume), and asks how multiple variables are organized within the individual and how this organization defines particular types, or categories, of people.

Calls for person-centered research have been made repeatedly for the past 50 years (e.g., Carlson, 1971). More recently, Pervin (1994) noted that trait researchers focus on individual differences rather than on the individuals themselves, and that "little attention is given to the question of pattern and organization," a "neglected area" of research (pp. 36–37). Until recently, the study of personality types has been held back by the lack of generally accepted procedures for deriving personality types empirically (see Robins, John, and Caspi, 1998, for a review). Thus, with the exception of Block's (1971) pioneering study, *Lives Through Time,* little systematic research was done on personality typology.

With the advent of the Big Five, however, researchers again became interested in studying the ways in which personality traits combine into coherent patterns within individuals and in identifying types of individuals that share the same basic personality profile. A series of recent studies has renewed the search for replicable personality types.

As shown in Table 4.6 (see next page), these studies varied greatly in the sex and age of the participants, their birth cohort and country of origin, as well as the type of data, instrument, and procedures used to derive the types. Nonetheless, three types recurred across all eight studies. In terms of their Big Five profiles, the type labeled Resilients showed a high level of adjustment and effective functioning on all five factors. In contrast, the types interpreted as Overcontrollers and Undercontrollers represent two different ways in which poor psychological adjustment can be manifested. The Overcontrollers had elevated scores on Agreeableness and Conscientiousness but scored low on Extraversion, whereas the Undercontrollers scored particularly low on Agreeableness and Conscientiousness and had elevated scores on Neuroticism.

Together, these studies demonstrate that replicable and generalizable personality types can be identified empirically. Validational studies further indicated that the unique constellation of traits associated with each type has important consequences for a wide range of life outcomes (Robins et al., 1998). These findings also suggest an integration of the Big Five dimensions with Block's (1971; Block & Block, 1980) dynamic conceptualization of personality functioning in terms of ego resilience and ego control. Block's dynamic constructs can be used to define the three replicable types, each of which captures a unique Big Five profile. More generally, the studies summarized in Table 4.6 show that the Big Five taxonomy is not only compatible with person-centered research but can help interpret personality types identified with different methods and in different cultures. Moreover, the Big Five need typological and dynamic elaboration if they are to fully account for personality structure. Conversely, person-centered typological research can make use of, and be informed by, the nomothetic Big Five dimensions, thus helping researchers develop dynamic accounts of personality functioning.

Theoretical Perspectives on the Big Five: Description and Explanation

Over the years, researchers have articulated a number of different perspectives on the conceptual status of the Big Five dimensions. Because the Big Five were first discovered in lexical research intended to provide a taxonomy of trait terms in the natural language, the factors were initially interpreted as dimensions of trait description or attribution (John et al., 1988). Subsequent research, however, has shown that the lexical factors converge with dimensions derived in other personality research traditions, that they have external or predictive validity (as reviewed above), and that all five of them show about equal amounts of heritability (Loehlin et al., 1998). Thus, it seems unlikely that these five dimensions are merely psycholexical artifacts or language phenomena. Given the evidence that

TABLE 4.6. Toward a Generalizable Personality Typology: Summary of Eight Studies Replicating Three Basic Types

	Studies in the United States			
	Block (1971)	York & John (1992)	Robins, John, Caspi, Moffitt, & Stouthamer-Loeber (1996)	Klohnen & Block (1996)
Personality types				
Type 1	Ego-Resilients	Individuated	Resilients	Resilients
Type 2	Vulnerable Overcontrollers	Traditional	Overcontrollers	Overcontrollers
Type 3	Unsettled Undercontrollers	Conflicted	Undercontrollers	Undercontrollers
Facets of generalizability				
Participants	84 boys/men	103 women	300 boys	106 men and women
Age	Both 13 and 35 years	43 years	12–13 years	23 years
Birth cohort	1920s	1937–1939	Late 1970s	1960s
Region	San Francisco area	San Francisco area	Pittsburgh	San Francisco area
Data source	Clinical judgments from data archives	Clinical judgments from data archives	Caregivers' reports	Interviewer's assessments
Instrument	Adult Q-set	Adult Q-set	Child Q-set	Adult Q-set
Type derivation	Q-factors across two time periods	Replicated Q-factors	Replicated Q-factors	Q-factors

	International replications			
	Caspi & Silva (1995)	van Lieshout, Haselager, Riksen-Walraven, & van Aken (1995)	Pulkkinen (1996)	Hart, Hofman, Edelstein, & Keller (1997)
Personality types				
Type 1	Well-adjusted	Resilients	Resilients/Individuated	Resilients
Type 2	Inhibited	Overcontrollers	Introverts/Anxious	Overcontrollers
Type 3	Undercontrolled	Undercontrollers	Conflicted/Under-controlled	Undercontrollers
Facets of generalizability				
Participants	1,024 boys and girls	79 boys and girls	275 men and women	168 boys and girls
Age	3 years	7, 10, and 12 years	26 years	7 years
Birth cohort	1972–1973	Early 1970s	1960s	1970
Region	New Zealand	The Netherlands	Finland	Iceland
Data source	Examiners' observations during a testing session	Teacher reports	Self-reports	Interviewer's assessments
Instrument	Behavior ratings	Child Q-set	Scale scores	Child Q-set
Type derivation	Replicated clusters	Cluster analysis	Cluster analysis	Replicated Q-factors

Note. Based on Robins et al. (1996) and Robins et al. (1998).

the Big Five dimensions refer to real individual differences, we need to ask how these differences should be conceptualized. A recent volume (Wiggins, 1996) addressed this issue, and we briefly summarize some of the major theoretical perspectives on the Big Five.

Researchers in the lexical tradition tend to take an agnostic stance regarding the conceptual status of traits. For example, Saucier and Goldberg (1996b) argued that their studies of personality description do not address issues of causality or the mechanisms underlying behavior. Their interest is primarily in the language of personality. This level of self-restraint may seem dissatisfactory to psychologists who are more interested in personality itself. However, the findings from the lexical approach are informative because the lexical hypothesis is essentially a functionalist argument about the trait concepts in the natural language. These concepts are of interest because language encodes the characteristics that are central, for cultural, social, or biological reasons, to human life and experience. Thus, Saucier and Goldberg argue that lexical studies define an agenda for personality psychologists because they highlight the important and meaningful psychological phenomena (i.e., phenotypic characteristics) that personality psychologists should study and explain. In other words, lexical researchers view issues such as the accuracy of self-descriptions and the causal origin of traits (i.e., genotypes) as open questions that need to be answered empirically. However, there may exist important characteristics that people may not be able to observe and describe verbally; if so, the agenda specified by the lexical approach may be incomplete and would need to be supplemented by more theoretically driven approaches (Block, 1995; Tellegen, 1993).

Several theories conceptualize the Big Five as relational constructs. In interpersonal theory (Wiggins & Trapnell, 1996), the theoretical emphasis is on the individual in relationships. The Big Five are taken to describe "the relatively enduring pattern of recurrent interpersonal situations that characterize a human life" (Sullivan, 1953, pp. 110–111), thus conceptualizing the Big Five as descriptive concepts. Wiggins and Trapnell emphasize the interpersonal motives of Agency and Communion, and interpret all of the Big Five dimensions in terms of their interpersonal implications. Because Extraversion and Agreeableness are the most clearly interpersonal dimensions in the Big Five, they receive conceptual priority in this model.

Socioanalytic theory (Hogan, 1996) focuses on the social functions of self- and other-perceptions. According to Hogan, trait concepts serve as the "linguistic tools of observers" (p. 172) used to encode and communicate reputations. This view implies that traits are socially constructed to serve interpersonal functions. Because trait terms are fundamentally about reputation, individuals who self-report their traits engage in a symbolic-interactionist process of introspection (i.e., the individual considers how others view him or her). Hogan emphasizes that individuals may distort their self-reports with self-presentational strategies; another source of distortion are self-deceptive biases (cf. Paulhus & John, 1998) which do not reflect deliberate impression management but honestly held, though biased, beliefs about the self.

The evolutionary perspective on the Big Five holds that humans have evolved "difference-detecting mechanisms" to perceive individual differences that are relevant to survival and reproduction (Buss, 1996, p. 185; see also Botwin, Buss, & Shackelford, 1997). Buss views personality as an "adaptive landscape" in which the Big Five traits represent the most salient and important dimensions of the individual's survival needs. The evolutionary perspective equally emphasizes person-perception and individual differences: Because people vary systematically along certain trait dimensions, and because knowledge of others' traits has adaptive value, humans have evolved a capacity to perceive those individual differences that are central to adaptation to the social landscape. The Big Five summarize these centrally important individual differences.

McCrae and Costa (1996; see also Chapter 5, this volume) view the Big Five as causal personality dispositions. Their five-factor theory (FFT) is an explanatory interpretation of the empirically derived Big Five taxonomy. The FFT is based on the finding that all of the Big Five dimensions have a substantial genetic basis (Loehlin et al., 1998) and must therefore derive, in part, from biological structures and processes, such as specific gene loci, brain regions (e.g., the amygdala), neurotransmitters (e.g., dopamine), hormones (e.g., testosterone), and so on (Plomin & Caspi, Chapter 9, this volume); it is in this sense that traits have causal status. McCrae and Costa distinguish between "basic tendencies" and "characteristic adaptations." Personality traits are basic tendencies that refer to the abstract underlying potentials of the individual; whereas attitudes, roles, relationships, and goals

are characteristic adaptations that reflect the interactions between basic tendencies and environmental demands accumulated over time. According to McCrae and Costa, basic tendencies remain stable across the life course, whereas characteristic adaptations can undergo considerable change. From this perspective, then, a statement such as "Paul likes to go to parties *because* he is extraverted" is not circular, as it would be if "extraverted" were merely a description of typical behavior (Wiggins, 1997). Instead, the concept "extraverted" stands in for biological structures and processes that remain to be discovered. This view is similar to Allport's (1937) account of traits as neuropsychic structures and Eysenck's view of traits as biological mechanisms (Eysenck & Eysenck, 1985).

The idea that personality traits have a biological basis is also fundamental to Gosling's (1999) proposal for a comparative approach to personality that studies individual differences in both human and nonhuman animals. Although scientists are reluctant to ascribe personality traits, emotions, and cognitions to animals, evolutionary theory predicts cross-species continuities not only for physical but also for behavioral traits; for example, Darwin (1998/1872) argued that emotions exist in both human and nonhuman animals. A recent review of 19 studies of personality factors in 12 nonhuman species showed substantial evidence for cross-species continuities (Gosling & John, 1999). Chimpanzees, various other primates, nonprimate mammals, and even guppies and octopuses all showed reliable individual differences in Extraversion and Neuroticism, and all but guppies and octopuses varied in Agreeableness as well, suggesting that these three Big Five factors may capture fundamental dimensions of individual differences across species. Further evidence suggests that elements of Openness (such as curiosity and playfulness) are present in at least some nonhuman animals. In contrast, only humans and our closest relatives, chimpanzees, appear to show systematic individual differences in Conscientiousness. Given the relatively complex social-cognitive functions involved in this dimension (i.e., following norms and rules, thinking before acting, and controlling impulses), it makes sense that Conscientiousness may have appeared rather recently in our evolutionary history. The careful application of ethological and experimental methodology and the high interobserver reliability in these studies make it unlikely that these findings reflect anthropomorphic projections. Rather, these surprising cross-species commonalities suggest that personality traits are caused, in part, by biological mechanisms that are shared by many species.

In conclusion, researchers hold a diversity of perspectives on the conceptual status of the Big Five, ranging from purely descriptive concepts to biologically based causal concepts. This diversity may seem to suggest that researchers cannot agree about the definition of the trait concept and that the field is in disarray (e.g., Pervin, 1994). It is important to recognize, however, that the various theoretical perspectives are not mutually exclusive. For example, although Saucier and Goldberg (1996b) caution against drawing inferences about genotypes from lexical studies, the lexical hypothesis does not preclude the possibility that the Big Five are embodied in biological structures and processes. In our view, "what is a trait?" is fundamentally an empirical question. Research in diverse areas such as behavior genetics (Plomin & Caspi, Chapter 9, this volume), molecular genetics (Lesch, Bengel, Heils, & Sabol, 1996), personality stability and change (Costa & McCrae, 1994; Helson & Stewart, 1994), and accuracy and bias in interpersonal perception (Kenny, 1994; Robins & John, 1997; see also Robins, Norem, & Cheek, Chapter 18, this volume) will be instrumental in building and refining a comprehensive theoretical account of the Big Five.

CONCLUSIONS AND IMPLICATIONS

At the beginning of this chapter, we argued that a personality taxonomy should provide a systematic framework for distinguishing, ordering, and naming types and characteristics of individuals. Ideally, that taxonomy would be built around principles that are causal and dynamic, exist at multiple levels of abstraction or hierarchy, and offer a standard nomenclature for scientists working in the field of personality. The Big Five taxonomy does not yet meet this high standard. In contrast to the biological taxonomies, the Big Five taxonomy provides descriptive concepts that still need to be explicated theoretically, and a nomenclature that is still rooted in vernacular English.

The Big Five structure has the advantage that everybody can understand the words that define the factors, and disagreements about their meanings can be reconciled by establishing their most common usage. Moreover, the natural language

is not biased in favor of any existing scientific conceptions; although the atheoretical nature of the Big Five dimensions makes them less appealing to some psychologists, it also makes them more palatable to researchers that reject dimensions cast in a theoretical mold different from their own. Whatever the inadequacies of the natural language for scientific systematics, broad dimensions inferred from folk usage are *not* a bad place to start a taxonomy. Even in animal taxonomy, as G. G. Simpson has pointed out, "the technical system evolved from the vernacular" (1961, pp. 12–13).

Obviously, a system that initially derives from the natural language does not need to reify such terms indefinitely. Indeed, several of the dimensions included among the Big Five, most notably Extraversion and Neuroticism, have been the target of various physiological and mechanistic explanations (Rothbart, 1991; see also Clark & Watson, Chapter 16, this volume). Similarly, Block and Block's (1980) notion of Ego Control might shed some light on the mechanisms underlying Conscientiousness and Extraversion. Tellegen's (1985) interpretation of Extraversion and Neuroticism as persistent dispositions toward thinking and behaving in ways that foster, respectively, positive and negative affective experiences promises to connect the Big Five with individual differences in affective functioning which, in turn, may be studied in more tightly controlled laboratory settings. In a sense, the Big Five differentiate domains of individual differences that have similar surface manifestations. However, the structures and processes underlying them have only begun to be explicated. Explication in explanatory and mechanistic terms will change the definition and assessment of the Big Five dimensions as we know them today.

As Allport concluded, "scalable dimensions are useful dimensions, and we hope that work will continue until we reach firmer agreement concerning their number and nature" (1958, p. 252). As Allport had hoped, the work on scalable dimensions has continued since, and researchers have now reached a firmer consensus about them: There are five replicable, broad dimensions of personality, and they may be summarized by the broad concepts of Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. In our view, the Big Five taxonomy is a major step ahead, a long-due extension and improvement over earlier factor systems that tended to compete with each other, rather than establish commonalities and convergences. The Big Five taxonomy captures, at a broad level of abstraction, the commonalities among most of the existing systems of personality description, and provides an integrative descriptive model for personality research.

ACKNOWLEDGMENTS

This chapter summarizes and updates previous reviews by John (1990) and John and Robins (1993, 1998). The preparation of this chapter was supported in part by Grant No. MH49255 from the National Institutes of Mental Health to Oliver P. John, and by an NSF Predoctoral Fellowship to Sanjay Srivastava. The support and resources provided by the Institute of Personality and Social Research are also gratefully acknowledged. We are grateful to Samuel D. Gosling, James J. Gross and Richard W. Robins for their comments and suggestions on a previous draft of this chapter.

NOTES

1. This historically important report, available only as an obscure Air Force technical report, was reprinted in a special issue on the Big Five in the *Journal of Personality* (Tupes & Christal, 1992).

2. Saucier (1994a) abbreviated the 100-item TDA to a set of 40 "mini-markers" to obtain an even shorter measure.

3. The other scale with a relatively lower reliability was the TDA Emotional Stability scale. In an attempt to measure the stable pole of his scale (which after all is called Emotional Stability), Goldberg (1992) included adjectives such as imperturbable, unexcitable, undemanding, unemotional, and unenvious as factor markers. Note that these adjectives are negations of emotionality, rather than affirmations of stability, and as such they were answered less reliably even in our verbally sophisticated sample, probably because these words are less familiar and more difficult to understand. More generally, the problem is that English has few adjectives denoting emotional stability, and those that do often fail to uniquely define the emotionally stable pole of Neuroticism (e.g., stable, calm, contented, and unemotional failed to load highly negatively on the Neuroticism factor in John, 1990, as shown here in Table 4.2). On the BFI, the problem of measuring the stable pole was solved through the use of phrases, such as "Is emotionally stable, not easily upset" and "Remains calm in tense situations," which provide sufficient context to clarify the attribute being measured.

4. The full matrix is available from the authors.

5. These values are lower-bound estimates, probably because the participants in the introductory psychology subject pool had little motivation to complete the instruments with utmost care. For example, Benet-Martinez and John (1998; Study 2) found somewhat higher mean alpha coefficients for both BFI (.85) and NEO-FFI (.82), as well as higher mean convergent validity correlations (.77). Similarly, a reanalysis of data from Gross and John (1998) showed a mean convergent validity correlation across all three instruments of .79, which is slightly higher than the .75 we found here. On the other hand, Goldberg (1992) reported much lower convergent validity correlations between his TDA scales and the longer NEO PI, averaging .61 compared to the .68 found here.

6. To test more directly whether this cross-loading is indeed due to the placement of warmth, we examined the three warmth-related items included in the BFI and the TDA. Interestingly, all three items had a stronger correlation with Agreeableness than with Extraversion on the NEO-FFI, and the total warmth scale formed from the three items correlated .59 with Agreeableness and .45 with Extraversion. When

APPENDIX: THE BIG FIVE INVENTORY (BFI)

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1. Disagree strongly
2. Disagree a little
3. Neither agree nor disagree
4. Agree a little
5. Agree strongly

I See Myself as Someone Who . . .

___1. Is talkative
___2. Tends to find fault with others
___3. Does a thorough job
___4. Is depressed, blue
___5. Is original, comes up with new ideas
___6. Is reserved
___7. Is helpful and unselfish with others
___8. Can be somewhat careless
___9. Is relaxed, handles stress well
___10. Is curious about many different things
___11. Is full of energy
___12. Starts quarrels with others
___13. Is a reliable worker
___14. Can be tense
___15. Is ingenious, a deep thinker
___16. Generates a lot of enthusiasm
___17. Has a forgiving nature
___18. Tends to be disorganized
___19. Worries a lot
___20. Has an active imagination
___21. Tends to be quiet
___22. Is generally trusting
___23. Tends to be lazy
___24. Is emotionally stable, not easily upset
___25. Is inventive
___26. Has an assertive personality
___27. Can be cold and aloof
___28. Perseveres until the task is finished
___29. Can be moody
___30. Values artistic, aesthetic experiences
___31. Is sometimes shy, inhibited
___32. Is considerate and kind to almost everyone
___33. Does things efficiently
___34. Remains calm in tense situations
___35. Prefers work that is routine
___36. Is outgoing, sociable
___37. Is sometimes rude to others
___38. Makes plans and follows through with them
___39. Gets nervous easily
___40. Likes to reflect, play with ideas
___41. Has few artistic interests
___42. Likes to cooperate with others
___43. Is easily distracted
___44. Is sophisticated in art, music, or literature

Please check: Did you write a number in front of each statement?

BFI scale scoring ("R" denotes reverse-scored items):

Extraversion: 1, 6R, 11, 16, 21R, 26, 31R, 36; Agreeableness: 2R, 7, 12R, 17, 22, 27R, 32, 37R, 42; Conscientiousness: 3, 8R, 13, 18R, 23R, 28, 33, 38, 43R; Neuroticism: 4, 9R, 14, 19, 24R, 29, 34R, 39; Openness: 5, 10, 15, 20, 25, 30, 35R, 40, 41R, 44

warmth was partialled out, the discriminant validity correlations between Extraversion on the NEO-FFI and Agreeableness on the BFI and TDA were reduced substantially, from .36 to .08 for the BFI and from .41 to .12 for the TDA. Even the correlation between Extraversion and Agreeableness on the NEO-FFI itself was reduced from .25 to −.02. These results are consistent with those from the CFA: reclassifying warmth as a facet of Agreeableness would reduce the overlap between Extraversion and Agreeableness, even within the NEO-FFI, and improve both convergent and discriminant validity.

REFERENCES

Adams, S. H., Cartwright, L. K., Ostrove, J. M., & Stewart, A. J., (1998). Psychological predictors of good health in three longitudinal samples of educated midlife women. *Health Psychology, 17,* 412–420.

Allen, B. P., & Potkay, C. R. (1981). On the arbitrary distinction between states and traits. *Journal of Personality and Social Psychology, 41,* 916–928.

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Holt.

Allport, G. W. (1958). What units shall we employ? In G. Lindzey (Ed.), *Assessment of human motives* (pp. 238–260). New York: Rinehart.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47,* No. 211.

Almagor, M., Tellegen, A., & Waller, N. G. (1995). The Big Seven model: A cross-cultural replication and further exploration of the basic dimensions of natural language trait descriptors. *Journal of Personality and Social Psychology, 69,* 300–307.

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

Bales, R. F. (1970). *Personality and interpersonal behavior.* New York: Holt, Rinehart, & Winston.

Bales, R. F., & Cohen, S. P. (1979). *SYMLOG: A system for the multiple level observation of groups.* New York: The Free Press.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26.

Barron, F. (1968). *Creativity and personal freedom.* Princeton, NJ: Van Nostrand.

Baumgarten, F. (1933). 'Die Charktereigenschaften.' [The character traits]. In *Beitraege zur Charakter- und Persoenlichkeitsforschung* (Whole No. 1). Bern, Switzerland: A. Francke.

Becker, W. C. (1960). The matching of behavior rating and questionnaire personality factors. *Psychological Bulletin, 57,* 201–212.

Benet-Martinez, V., & John, O. P. (1998). *Los Cinco Grandes* across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75,* 729–750.

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.

Bentler, P. E. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research* (Reprint Ed., 1978). Palo Alto, CA: Consulting Psychologists Press.

Block, J. (1971). *Lives through time.* Berkeley, CA: Bancroft Books.

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

Block, J., & Block, J. H. (1969). *The California Child Q-set.* Palo Alto, CA: Consulting Psychologists Press.

Block, J. H., & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior. In W. A. Collins (Ed.), *Minnesota symposia on child psychology* (Vol. 13, pp. 39–101). Hillsdale, NJ: Erlbaum.

Bond, M. H. (1979). Dimensions used in perceiving peers: Cross-cultural comparisons of Hong Kong, Japanese, American and Filipino university students. *International Journal of Psychology, 14,* 47–56.

Bond, M. H. (1983). Linking person perception dimensions to behavioral intention dimensions: The Chinese connection. *Journal of Cross-Cultural Psychology, 14,* 41–63.

Bond, M. H., & Forgas, J. P. (1984). Linking person perception to behavior intention across cultures: The role of cultural collectivism. *Journal of Cross-Cultural Psychology, 15,* 337–352.

Bond, M. H., Nakazato, H., & Shiraishi, D. (1975). Universality and distinctiveness in dimensions of Japanese person perception. *Journal of Cross-Cultural Psychology, 6,* 346–357.

Borgatta, E. F. (1964). The structure of personality characteristics. *Behavioral Science, 9,* 8–17.

Botwin, M. D., & Buss, D. M. (1989). Structure of act-report data: Is the five-factor model of personality recaptured? *Journal of Personality and Social Psychology, 56,* 988–1001.

Botwin, M. D., Buss, D. M., & Shackelford, T. K. (1997). Personality and mate preferences: Five factors in mate selection and marital satisfaction. *Journal of Personality, 65,* 107–136.

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

Burisch, M. (1984). Approaches to personality inventory construction. *American Psychologist, 39,* 214–227.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development.* New York: Wiley.

Buss, D. M. (1996). Social adaptation and five major factors of personality. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 180–207). New York: Guilford Press.

Caprara, G. V., & Perugini, M. (1994). Personality described by adjectives: The generalizability of the Big Five to the Italian lexical context. *European Journal of Personality, 8,* 357–369.

Carlson, R. (1971). Where is the person in personality research? *Psychological Bulletin, 75,* 203–219.

Caspi, A. (1998). Personality development across the life course. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Social, emotional, and personality development* (5th ed., pp. 311–388). New York: Wiley.

Caspi, A., & Silva, P. A. (1995). Temperamental qualities at age 3 predict personality traits in young adulthood: Longitudinal evidence from a birth cohort. *Child Development, 66,* 486–498.

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. (1945a). The description of personality: Principles and findings in a factor analysis. *American Journal of Psychology, 58,* 69–90.

Cattell, R. B. (1945b). The principal trait clusters for describing personality. *Psychological Bulletin, 42,* 129–161.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF).* Champaign, IL: IPAT.

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology, 54,* 541–557.

Church, T. A., & Katigbak, M. S. (1989). Internal, external, and self-report structure of personality in a non-western culture: An investigation of cross-language and cross-cultural generalizability. *Journal of Personality and Social Psychology, 57,* 857–872.

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.

Comrey, A. L. (1970). *Manual for the Comrey Personality Scales.* San Diego, CA: Educational and Industrial Testing Service.

Conley, J. J. (1985). Longitudinal stability of personality traits: A multitrait–multimethod–multioccasion analysis. *Journal of Personality and Social Psychology, 49,* 1266–1282.

Costa, P. T., & McCrae, R. R. (1976). Age differences in personality structure: A cluster analytic approach. *Journal of Gerontology, 31,* 564–570.

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64,* 21–50.

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1989). *NEO PI/FFI manual supplement.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R Professional manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1994). Set like plaster: Evidence for the stability of adult personality. In T. F. Heatherton & J. L Weinberger (Ed.), *Can personality change?* (pp. 21–40). Washington, DC: American Psychological Association.

Costa, P. T., & Widiger, T. A. (Eds.). (1994). *Personality disorders and the five-factor model of personality.* Washington, DC: American Psychological Association.

Darwin, C. (1998). *The expression of emotions in man and animals,* 3rd ed., P. Ekman, Ed. New York: Oxford. (Original work published 1872.)

De Raad, B. (1994). An expedition in search of a fifth universal factor: Key issues in the lexical approach. *European Journal of Personality, 8,* 229–250.

De Raad, B., Di Blas, L., & Perugini, M. (1998). Two independently constructed Italian trait taxonomies: Comparisons among Italian and between Italian and Germanic languages. *European Journal of Personality, 12,* 19–41.

De Raad, B., Mulder, E., Kloosterman, K., & Hofstee, W. K. (1988). Personality-descriptive verbs. *European Journal of Personality, 2,* 81–96.

De Raad, B., Perugini, M., Hrebickova, M., & Szarota, P. (1998). Lingua franca of personality: Taxonomies and structures based on the psycholexical approach. *Journal of Cross-Cultural Psychology, 29,* 212–232.

Digman, J. M. (1963). Principal dimensions of child personality as seen in teachers' judgments. *Child Development, 34,* 43–60.

Digman, J. M. (1989). Five robust trait dimensions: Development, stability, and utility. Special Issue: Long-term stability and change in personality. *Journal of Personality, 57,* 195–214.

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., & Shmelyov, A. G. (1996). The structure of temperament and personality in Russian children. *Journal of Personality and Social Psychology, 71,* 341–351.

Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis and comparison of six major studies. *Multivariate Behavioral Research, 16,* 149–170.

Donahue, E. M. (1994). Do children use the Big Five, too? Content and structural form in personality description. *Journal of Personality, 62,* 45–66.

Eysenck, H. J. (1986). Models and paradigms in personality research. In A. Angleitner, A. Furnham, & G. Van Heck (Eds.) *Personality psychology in Europe, Vol. 2: Current Trends and Controversies* (pp. 213–223). Lisse, The Netherlands: Swets & Zeitlinger.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5, or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences, 12,* 773–790.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences, 13,* 667–673.

Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology, 73,* 1224–1237.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.* New York: Plenum Press.

Field, D., & Millsap, R. E. (1991). Personality in advanced old age: Continuity or change? *Journals of Gerontology, 46,* 299–308.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology, 44,* 329–344.

Friedman, H. S., Hawley, P. H., & Tucker, J. S. (1994). Personality, health, and longevity. *Current Directions in Psychological Science, 3,* 37–41.

Friedman, H. S., Tucker, J. S., Schwartz, J. E., & Tomlinson-Keasey, C. (1995). Psychosocial and behavioral predictors of longevity: The aging and death of the "Termites." *American Psychologist, 50,* 69–78.

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.

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

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., & Rosolack, T. K. (1994). The Big Five factor structure as an integrative framework: An empirical comparison with Eysenck's P-E-N model. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood.* Hillsdale, NJ: Erlbaum.

Gosling, S. D. (1999). *From mice to men: What can animal research tell us about personality?* Manuscript submitted for publication.

Gosling, S. D., & John, O. P. (1999). Personality dimensions in non-human animals: A cross-species review. *Current Directions in Psychological Science, 8,* 69–75.

Gough, H. G. (1979). A creative personality scale for the Adjective Check List. *Journal of Personality and Social Psychology, 37,* 1938–1405.

Gough, H. G. (1987). *The California Psychological Inventory administrator's guide.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G., & Heilbrun, A. B., Jr. (1983). *The Adjective Check List manual.* Palo Alto, CA: Consulting Psychologists Press.

Gould, S. J. (1981). *The mismeasure of man.* New York: Norton.

Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 795–824). San Diego, CA: Academic Press.

Graziano, W. G., & Ward, D. (1992). Probing the Big Five in adolescence: Personality and adjustment during a developmental transition. *Journal of Personality, 60,* 425–439.

Gross, J. J., & John, O. P. (1998). Mapping the domain of expressivity: Multimethod evidence for a hierarchical model. *Journal of Personality and Social Psychology, 74,* 171–190.

Guilford, J. P. (1974). Rotation problems in factor analysis. *Psychological Bulletin, 81,* 498–501.

Guilford, J. P. (1975). Factors and factors of personality. *Psychological Bulletin, 82,* 802–814.

Guthrie, G. M., & Bennett, A. B., Jr. (1971). Cultural differences in implicit personality theory. *International Journal of Psychology, 6,* 305–312.

Halverson, C. F., Kohnstamm, G. A., & Martin R. P. (1994). *The developing structure of temperament and personality from infancy to adulthood.* Hillsdale, NJ: Erlbaum.

Hampson, S. E., Goldberg, L. R., & John, O. P. (1987). Category-breadth and social-desirability values for 573 personality terms. *European Journal of Personality, 1,* 241–258.

Hampson, S. E., John, O. P., & Goldberg, L. R. (1986). Category breadth and hierarchical structure in personality: Studies of asymmetries in judgments of trait implications. *Journal of Personality and Social Psychology, 51,* 37–54.

Hart, D., Hofman, V., Edelstein, W., & Keller, M. (1997). The relation of childhood personality types to adolescent behavior and development: A longitudinal study of Icelandic children. *Developmental Psychology, 33,* 195–205

Heatherton, T. F., & Weinberger, J. L. (Eds.). (1994). *Can personality change?* Washington, DC: American Psychological Association.

Helson, R. (1967). Personality characteristics and developmental history of creative college women. *Genetic Psychology Monographs, 76,* 205–256.

Helson, R. (1985). Which of those young women with creative potential became productive? Personality in college and characteristics of parents. In R. Hogan and W. H. Jones (Eds.), *Perspectives in personality theory, measurement, and interpersonal dynamics* (Vol. 1, pp. 49–80). Greenwich, CT: JAI Press.

Helson, R., & Stewart, A. (1994). Personality change in adulthood. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 21–40). Washington, DC: American Psychological Association.

Hofstee, W. K., 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., Kiers, H. A., 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.

Hogan, R. (1983). A socioanalytic theory of personality. In M. Page (Ed.) *Nebraska symposium on motivation, 1982: Personality—Current theory and research.* Lincoln, NE: University of Nebraska Press.

Hogan, J., & Ones, D. S. (1997). Conscientiousness and integrity at work. In R. Hogan, J. A. Johnson, and S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 849–870). San Diego, CA: Academic Press.

Hogan, R. (1986). *Hogan Personality Inventory manual.* Minneapolis, MN: National Computer Systems.

Hogan, R. (1996). A socioanalytic perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 180–207). New York: Guilford Press.

Hrebickova, M., & Ostendorf, F. (1995). Lexikalni pristup k osobnosti. V: Klasifikace pridavnych jmen do kategorii osobnostni deskripce. [Lexical approach to personality: V. Classification of adjectives into categories of personality description]. *Ceskoslovenska Psychologie, 39,* 265–276.

Huey, S.J., & Weisz, J. R. (1997). Ego control, ego resiliency, and the five-factor model as predictors of behavioral and emotional problems in clinic-referred children and adolescents. *Journal of Abnormal Psychology, 106,* 404–415.

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

John, O. P. (1989). Towards a taxonomy of personality descriptors. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 261–271). New York: Springer.

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., 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., & Benet-Martinez, V. (in press). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology.* Cambridge, UK: Cambridge University Press.

John, O. P., Caspi, A., Robins, R. W., Moffitt, T. E., & Stouthamer-Loeber, M. (1994). The "Little Five": Exploring the nomological network of the five-factor model of personality in adolescent boys. *Child Development, 65,* 160–178.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

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, and N. Smid (Eds.), *Personality psychology in Europe: Theoretical and empirical developments* (pp. 83–100). Berwyn, PA: Swets North America Inc.

John, O. P., & Gosling, S. D. (in press). Personality traits. In A. E. Kazdin (Ed.), *Encyclopedia of psychology.* Washington, DC: American Psychological Association.

John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). Is there a basic level of personality description? *Journal of Personality and Social Psychology, 60,* 348–361.

John, O. P., & Robins, R. W. (1993). Gordon Allport: Father and critic of the five-factor model. In K. H. Craik, R. T. Hogan, & R. N. Wolfe (Eds.), *Fifty years of personality research* (pp. 215–236). New York: Plenum Press.

John, O. P., & Robins, R. W. (1998). Recent trends in Big Five research: Development, predictive validity, and personality types. In J. Bermudez et al. (Eds.), *Personality Psychology in Europe* (Vol. 6, pp. 6–16). Tilburg, The Netherlands: Tilburg University Press.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford Press.

Klages, L. (1926). *The science of character* (Translated 1932). London: George Allen and Unwin. (Original work published 1926.)

Klohnen, E. C., & Block, J. (1996). Unpublished data, Department of Psychology, Berkeley, California.

Kohnstamm, G. A., Halverson, C. F., Mervielde, I., & Havill, V. L. (Eds.). (1998). *Parental descriptions of child personality: Developmental antecedents of the Big Five?* Mahwah, NJ: Erlbaum.

Lesch, K. P., Bengel, D., Heils, A., & Sabol, S. Z. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science, 274,* 1527–1531.

Loehlin, J. C., McCrae, R. R., Costa, P. T., Jr., & John, O. P. (1998). Heritabilities of common and measure-specific components of the Big Five personality factors. *Journal of Research in Personality, 32,* 431–453.

MacKinnon, D. W. (1965). Personality and the realization of creative potential. *American Psychologist, 20,* 273–281.

McAdams, D. P. (1992). The five-factor model in personality: A critical appraisal. *Journal of Personality, 60,* 329–361.

McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality, 63,* 365–396.

McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin, 120,* 323–337.

McCrae, R. R., & Costa, P. T. (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. (1985b). Openness to experience. In R. Hogan & W. H. Jones, *Perspectives in personality* (Vol. 1, pp. 145–172). Greenwich, CT: JAI Press.

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

McCrae, R. R., & Costa, P. T. (1990). *Personality in adulthood.* New York: Guilford Press.

McCrae, R. R., & Costa, P. T. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five- factor model of personality: Theoretical perspectives* (pp. 51–87). New York: Guilford Press.

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509–516.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60,* 175–215.

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.

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11,* 145–165.

Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers–Briggs Type Indicator.* Palo Alto, CA: Consulting Psychologists Press.

Nakazato, H., Bond, M. H., & Shiraishi, D. (1976). Dimensions of personality perception: An examination of Norman's hypothesis. *Japanese Journal of Psychology, 47,* 139–148.

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.

Norman, W. T. (1967). *2,800 personality trait descriptors: Normative operating characteristics for a university population.* Department of Psychology, University of Michigan, Ann Arbor, MI.

Nowakowska, M. (1973). The limitations of the factor-analytic approach to psychology with special application to Cattell's research strategy. *Theory and Decision, 4,* 109–139.

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

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66,* 1025–1060.

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

Pervin, L. A. (1994). A critical analysis of current trait theory. *Psychological Inquiry,* 5, 103–113.

Pervin, L. A., & John, O. P. (1997). *Personality: Theory and research* (7th ed.). New York: Wiley.

Pulkkinen, L. (1996). Female and male personality styles: A typological and developmental analysis. *Journal of Personality and Social Psychology, 70,* 1288–1306.

Revelle, W. (1987). Personality and motivation: Sources of inefficiency in cognitive performance. *Journal of Research in Personality, 21,* 436–452.

Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. Hogan, J. A. Johnson, & S. R. Briggs (Ed.), *Handbook of personality psychology* (pp. 649–679). San Diego, CA: Academic Press.

Robins, R. W., John, O. P., & Caspi, A. (1994). Major dimensions of personality in early adolescence: The Big Five and beyond. In C. F. Halverson, J. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 267–291). Hillsdale, NJ: Erlbaum.

Robins, R. W., John, O. P., & Caspi, A. (1998). The typological approach to studying personality. In R. B. Cairns, J. Kagan, & L. Bergman (Eds.), *The individual in developmental research: Essays in honor of Marian Radke-Yarrow* (pp. 135–160). Beverly Hills, CA: Sage

Robins, R. W., John, O. P., Caspi, A., Moffitt, T. E., & Stouthamer-Loeber, M. (1996). Resilient, overcontrolled, and undercontrolled boys: Three replicable personality types? *Journal of Personality and Social Psychology, 70,* 157–171.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rothbart, M. (1991). Temperament and development. In G. A. Kohnstamm, L. Bates, and M. K. Rothbart (Eds.), *Handbook of temperament in childhood.* Sussex, UK: Wiley.

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

Saucier, G. (1994a). Mini-Markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment, 63,* 506–516.

Saucier, G. (1994b). Trapnell versus the lexical factor: More ado about nothing? *European Journal of Personality, 8,* 291–298.

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., & Goldberg, L. R. (1996a). Evidence for the Big Five 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.

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

Simpson, G. G. (1961). *Principles of animal taxonomy.* New York: Columbia University Press.

Somer, O., & Goldberg, L. R. (1999). The structure of Turkish trait-descriptive adjectives. *Journal of Personality and Social Psychology, 76,* 431–450.

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry.* New York: Norton.

Szarota, P. (1995). Polska Lista Przymiotnikowa (PLP): Narzedzie do diagnozy Pieciu Wielkich czynnikow osobowosci [Polish Adjective List: An instrument to assess the five-factor model of personality]. *Studia Psychologiczne, 33,* 227–256.

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

Tellegen, A. (1982). *Brief manual for the Differential Personality Questionnaire.* Unpublished manuscript, University of Minnesota.

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.

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

Trapnell, P. D., & Wiggins, J. S. (1990). Extension of the Interpersonal Adjective scales to include the Big Five dimensions of personality. *Journal of Personality and Social Psychology, 59,* 781–790.

Tupes, E. C., & Christal, R. C. (1961). *Recurrent personality factors based on trait ratings* (Tech. Rep.). Lackland Air Force Base, TX: USAF.

Tupes, E. C., & Christal, R. C. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality, 60,* 225–251.

van Lieshout, C. F. M., & Haselager, G. J. T. (1994). The Big Five personality factors in Q-sort descriptions of children and adolescents. In C. F. Halverson, G. A. Kohnstamm, and R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 293–318). Hillsdale, NJ: Erlbaum.

van Lieshout, C. F. M., Haselager, G. J. T., Riksen-Walraven, J. M., & van Aken, M. A. (1995, April). *Personality development in middle childhood.* Paper presented at the Society for Research in Child Development, Indianapolis, Indiana.

Watson, D., & Clark, L. A. (1997). Extraversion and its positive emotional core. In R. Hogan, J. A. Johnson, and S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 767–793). San Diego, CA: Academic Press.

White, G. M. (1980). Conceptual universals in interpersonal language. *American Anthropologist, 82,* 759–781.

Wiggins, J. S. (1968). Personality structure. In P. R. Farnsworth (Ed.), *Annual review of psychology* (Vol. 19, pp. 320–322). Palo Alto: Annual Reviews.

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. (1995). *Interpersonal Adjective Scales: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Wiggins, J. S. (1997). In defense of traits. In R. Hogan, J. A. Johnson, & S. R. Briggs (Ed.), *Handbook of personality psychology* (pp. 649–679). San Diego, CA: Academic Press.

Wiggins, J. S. (Ed.). (1996). *The five-factor model of personality: Theoretical perspectives.* New York: Guilford Press.

Wiggins, J. S., & Pincus, A. L. (1989). Conceptions of personality disorders and dimensions of personality. *Psychological Assessment, 1,* 305–316.

Wiggins, J. S., & Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 180–207). New York: Guilford Press.

Yang, K.-S., & Bond, M. H. (1990). Exploring implicit personality theories with indigenous or imported constructs: The Chinese case. *Journal of Personality and Social Psychology, 58,* 1087–1095.

Yik, M. S., & Bond, M. H. (1993). Exploring the dimensions of Chinese person perception with indigenous and imported constructs: Creating a culturally balanced scale. *International Journal of Psychology, 28,* 75–95.

York, K. L., & John, O. P. (1992). The four faces of Eve: A typological analysis of women's personality at midlife. *Journal of Personality and Social Psychology, 62,* 494–508.