Journal of Personality and Social Psychology
1997, Vol. 73, No. 6, 1296–1312

0022-3514/97/$3.00

Effects of Variable Selection on the Factor Structure of Person Descriptors

Gerard Saucier
California State University, San Bernardino

Previous factor-analytic studies of lexical person descriptors have produced some recurrent patterns of results, but their integration has been hampered by divergences in variable sampling, such as disparate criteria for what is considered a personality descriptor. To isolate effects of variable selection on factor structures, 500 of the most familiar English person descriptors were identified. Fifteen judges provided reliable classifications of these adjectives as disposition, state, social evaluation, or physical–appearance terms. Analyses of adult self-ratings (N = 700) and acquaintance ratings (N = 201) led to a stable Big Five structure when disposition terms, or combined disposition and state terms, were analyzed. Including a wider range of terms led to two additional stable factors: Attractiveness and a factor resembling Big Seven Negative Valence. A stable 3-factor solution was relatively impervious to variable-selection effects.

A prime task of the science of psychology is the identification of the most important ways in which people vary. Factor analysis can help reduce the many aspects of variation to a few basic dimensions. Evidence regarding the basic dimensions of personality variation can have major implications for both measures and theories.

An increasing number of researchers have adopted the view that phenotypic personality variation is most meaningfully distinguished in terms of five broad orthogonal factors, labeled the Big Five (Digman, 1990; Goldberg, 1990, 1993; John, 1990; Ostendorf, 1990a) or the five-factor model (McCrae & Costa, 1987). Although the number five may currently command the most respect, alternative schools of thought champion the cause of other "magic numbers." Eysenck (1970, 1992) advocated three factors; Hogan (1986) suggested six; Tellegen (1993) proposed seven. Consensual resolution of these disputes on the basis of evidence has proven difficult.

One source of these disputes is differences among investigations in their preferred range of variable selection. As J. Block (1995) noted, changing the set of variables used in a factor analysis can lead to a change in the resulting factor structure. Just as personality can be defined in more than one way (Allport, 1937), so can the criteria for selecting personality-relevant descriptor variables. Thus, Allport and Odbert (1936; also Cattell, 1943) differentiated descriptors of temporary states and social evaluations from those trait descriptors that they deemed most personality relevant. Eysenck (1970) emphasized biologically based disposition variables, excluding intelligence and abilities on the one hand and attitudes on the other. Other investigators have cast wider nets. Studies of German personality descriptors (Angleitner, Ostendorf, & John, 1990; Ostendorf, 1990a) included temperament and ability terms together. Goldberg (1990, 1992) included some mood and attitude terms (e.g., Anxious, Jealous, Conservative) that can also refer to dispositions. Tellegen and Waller (1987) included not only moods and attitudes but also social evaluation terms (e.g., *awful, excellent*). Because they used different criteria for variable selection, it is little surprise that these studies failed to all converge on a single structure. This report focuses on a search for systematic and predictable patterns in the effects of variable selection on the factor structure of person descriptors.

Toward Optimal Variable Sampling in Personality Psychology

From where do personality variables come? Personality questionnaires are usually composed of phrases or sentences. Because the "population" of possible questionnaire items is essentially infinite, it would be difficult to infer that any set of such items were a representative sample of all possible phrases and sentences. A firmer basis for representative variable selection requires recourse to a finite population of variables. Commonly used single words in a language are such a population.

Therefore, a useful beginning point for scientific classifications of personality attributes is the natural language (Goldberg, 1981; Tellegen, 1993). Personality attributes are observable and socially meaningful phenomena, and developing concepts to summarize these phenomena has been a human occupation for many thousands of years; thus, high overlap between lay and

Gerard Saucier, Department of Psychology, California State University, San Bernardino.

This work was supported by National Institute of Mental Health Grant MH-49227. Terry Maul, Susan Shodahl, Rebecca Deighan, and Tom Kipnis assisted with the recruitment of acquaintance rating participants. Lewis R. Goldberg assisted with the recruitment of community resident participants for the familiarity rating and self-report tasks, and provided helpful editorial suggestions. For feedback regarding earlier versions of this article, I am grateful to Veronica Benet, A. Timothy Church, W. K. B. Hofstee, Fritz Ostendorf, Dean Peabody, and Auke Tellegen.

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

scientific descriptive concepts is to be expected (Saucier & Goldberg, 1996b). The lexical approach can produce a descriptive nonexplanatory framework (Greenwood, 1991) that provides minimum criteria for representative variable sampling: It can identify features too important to leave out (Goldberg & Saucier, 1995). The content validity of numerous personality inventories (e.g., Cattell, Eber, & Tatsuoka, 1970; Costa & McCrae, 1992; Hogan, 1986; Wiggins, Trapnell, & Phillips, 1988) is, in part, rooted in the lexical approach.

For the purpose of studying human attributes, the most useful subset of single words is the set of adjectives that can be used to denote human propensities and qualities (Saucier & Goldberg, 1996b). Adjectives (e.g., *cynical, kind*) are advantageous because they richly represent both desirable and undesirable attributes applied to self and others, whereas personality type-nouns (e.g., *cynic, jerk*) represent in large part undesirable attributes used in reacting to and labeling others (cf. Wierzbicka, 1986). Among adjectives that denote attributes, frequently used terms are of more importance in everyday transactions than are unfamiliar terms. This rationale forms part of the lexical hypothesis, which states that (a) the most distinctive, significant, and widespread phenotypic attributes tend to become encoded as single words in the conceptual reservoir of language, and (b) the degree of representation of an attribute in language tends to correspond with the relative importance of the attribute (Saucier & Goldberg, 1996b; cf. Osgood, May, & Miron, 1975, p. 45). Accordingly, searching a dictionary for person-descriptive adjectives is a useful variable-sampling strategy, especially if the terms retained are those that are most frequently used.

However, within the broad lexical strategy, the tactics of variable sampling differ from one study to another. A useful reference point for comparisons is the influential four-column classification scheme introduced by Allport and Odbert (1936). Stable traits (e.g., *rebellious, daring*) were distinguished from social evaluations, "terms descriptive of present activity, temporary states of mind, and mood" (p. 26), and miscellaneous terms, including those referring to abilities and appearance. Social evaluations (e.g., *exciting, desirable, exceptional*) were deemed less relevant to the study of personality, because they refer to "value estimates" and to "social stimulus value"—the individual's effect on others or degree of approval by others. Terms for temporary states (e.g., *sad, excited*) were excluded because they were not "enduring and recurring modes of adjustment" (Allport & Odbert, 1936, p. 26). Lexical studies have differed in their handling of states and social evaluations. With no clear consensus as to which attributes ought to be counted as personality variables, the most informative procedure would be to sample broadly from attributes of diverse types and carry out a reliable classification of the descriptors into variable categories, thus enabling some control for the effects of variable selection.

How large a sample of variables ought to be taken from the natural language? The most stable (reliable, robust, replicable, and generalizable) factor structures are likely to come from analyses featuring not only large samples of participants but also a large ratio of variables to factors (Guadagnoli & Velicer, 1988; Velicer & Fava, 1987). It is preferable, then, that several hundred person descriptors be selected.

Cross-cultural generalizations about personality attributes require studies in different cultures and different languages. The most consequential factors are those found in independent within-language (emic) studies. A demanding test is not provided by the importation of a selection of (etic) variables from one language to another, to see whether it holds its expected structure. Far stronger evidence for robust factors is obtained when (a) the most important person descriptors within a new language are identified, (b) the factor structure of person descriptors among a representative sample of users of that language is examined, and (c) this structure is compared to structures derived "indigenously" within other languages. Useful evidence as to the basic dimensional structure for personality attributes can be drawn from an aggregate of studies, with representative variable sampling, conducted within diverse languages and cultures (Goldberg, 1981).

Patterns in Lexical Factor Studies of Person Descriptors

Over the last decade, such an aggregation of studies has begun to accumulate. Studies featuring indigenously derived personality-relevant adjectives in the description of actual target persons have now been conducted in nine distinct languages. These studies can be sorted into three groups, based on variable-sampling tactics and on whether five or seven factors were judged to be optimal for comprehending the data.

Studies of Disposition Descriptors

One group of studies includes those conducted in Dutch (De Raad, Hendriks, & Hofstee, 1992; also Brokken, 1978; Hofstee, Brokken, & Land, 1981), Italian (Caprara & Perugini, 1994), and Hungarian (Szirmak & De Raad, 1994). Variable selection in these studies approximated most closely that of Allport and Odbert (1936) with an emphasis on stable traits only. In each study, judges provided utility or personality-relevance ratings (e.g., fit to sentence stems like *X is a* _____ *person* and *X is* _____ *by nature*); those terms having the highest mean ratings were retained. Retained terms apparently included relatively few referring to abilities or talents. Each of these three studies found rather clear replicas of the first four of the Big Five: Extraversion, Agreeableness, Conscientiousness, and Emotional Stability. With regard to the fifth factor, agreement was not as strong. The Dutch factor emphasized attributes like critical and rebellious and the Italian factor emphasized unconventionality. The Hungarian counterpart was labeled Integrity, although the rotation of six rather than five factors led to an additional factor interpretable as Big Five Intellect (De Raad & Szirmak, 1994).

A second group of studies includes those conducted in German (Ostendorf, 1990a; Ostendorf & Angleitner, 1993), Czech (Hrebickova, Ostendorf, & Angleitner, 1995), and English (Saucier & Goldberg, 1996a; cf. Goldberg, 1990, 1992). These studies can be distinguished from the first group by a greater inclusion of terms referring to abilities and talents, which Allport and Odbert (1936) had relegated to the "miscellaneous" category. This effect was achieved in the German and Czech studies by pooling adjectives that judges had consensually classified as "temperament and character traits" with those classified as "abilities and talents, or their absence." In the genealogy

of the English-language studies, a similar effect was achieved when Norman (1967), in an influential classification extending the work of Allport and Odbert (1936), included many ability descriptors (e.g., *intelligent, creative*) in the "stable traits" category. The German, Czech, and English studies produced fairly similar Big Five factor structures, including a fifth factor clearly interpretable as Intellect.

All studies in both of these two groups were based on comprehensive catalogs of person-descriptive adjectives in the dictionaries of their six respective languages. As in Allport and Odbert (1936), most terms classifiable as either social evaluations or temporary states were excluded. Moreover, when the selected variables (samples of 243–561 adjectives, with infrequent terms generally excluded) were analyzed in data based on large samples (samples of 274–899 descriptions), each study found the Big Five, with cross-language divergence most apparent for the fifth factor. These convergent results suggest the following generalization: When adjectives referring to dispositions are gathered from any European language, and when a large sample of participants uses a large, representative sample of these adjectives in self- or peer descriptions, the adjectives will have a five-factor structure corresponding roughly to the Big Five model (Goldberg, 1993).

Studies of a Less Restricted Range of Descriptors

A third group of studies used an alternative set of tactics for variable selection, beginning with a page sampling of a dictionary. Terms referring to social evaluations and temporary states were not systematically excluded. The prototype study was that of Tellegen and Waller (1987; summarized also by Waller & Zavala, 1993). An abridged dictionary (*The American Heritage Dictionary,* 1982) was divided into roughly sixty 25-page sections, and 7 or 8 noncontiguous pages from each section were randomly selected. On each selected page, the first personality-descriptive adjective that could be fit into the stems *tends to be* _____ and *is often* _____ was extracted. The 400 extracted terms included social evaluations and state terms, as well as traits, and were used as stimuli for self-reports by 585 university students. Because many terms were unfamiliar, the items included phrases drawn from the dictionary definitions of the terms. Factor analyses (5–20 factors) were conducted; a seven-factor solution was judged the "most compelling and psychologically meaningful" (Waller, in press). The seven factors were deemed to correspond to the Big Five, plus Positive Valence (e.g., *important, outstanding*) and Negative Valence (e.g., *evil, vicious*). The last two factors drew on the social evaluation descriptors excluded by Allport and Odbert (1936) and all previous lexical factor studies.

The basic assumptions of a Big Seven model (Almagor, Tellegen, & Waller, 1995; Tellegen, 1993; Waller & Zavala, 1993) can be set out as follows: When ratings using an unrestricted range of person descriptors are factor analyzed, the best solution will have seven factors, the first five corresponding to those from the five-factor model (Costa & McCrae, 1992)—Positive Emotions (Extraversion), Agreeableness, Conscientiousness, Negative Emotions (Neuroticism; a reversal of Emotional Stability), and Conventionality (low Openness)—with the other two being evaluative or valence factors, one positive, one negative.

Variants of the prototype study were conducted with Spanish (Benet & Waller, 1995) and Hebrew (Almagor et al., 1995) person descriptors. In the Spanish study, the indigenous seven-factor solution clearly reproduced the first three of the lexical Big Five factors, as well as Positive Valence, these four correlating .60–.79 with corresponding factors from a Big Seven measure (Inventory of Personality Characteristics #7; Tellegen, Grove, & Waller, 1991). A Negative Valence factor was also present, although its correlation with the corresponding Big Seven factor was noticeably lower (.47). The other three factors were smaller, of lower saturation, and correlated much less with corresponding Big Seven factors.

In the Hebrew study, there was some correspondence between the indigenous seven-factor solution and the Big Seven (Tellegen & Waller, 1987), by a two-out-of-three consensus of Tellegen and Waller (using an English translation) and a third judge fluent in Hebrew. But this correspondence was quite modest; for each of the seven empirically derived Hebrew factors, fewer than 50% of the associated descriptors were categorized into the domain of any one of the a priori Big Seven factors. Scrutiny of the factor loadings (Almagor et al., 1995, p. 303) suggests that the factor labeled Positive Valence might better be called Intellect; moreover, the factor labeled Negative Valence, unlike its English and Spanish counterparts, was a clearly bipolar factor, virtues at one pole, vices at the other. There was no Conventionality factor evident in the Hebrew study. A statement that "findings attest to the robustness of the Big Seven model" (Almagor et al., 1995, p. 306) thus seems questionable.

This third group of studies showed that when lexical studies using large samples of descriptors and of participants use a wider variable selection, including social evaluation and temporary state terms, the resultant factor structure does not fall entirely into the Big Five pattern, not even when exactly five factors are rotated. Factors resembling the first three factors of the Big Five—Extraversion, Agreeableness, and Conscientiousness—did appear in consistent form. Emotional Stability, Positive Valence, and Negative Valence were each clearly apparent in two of the three studies. The seventh factor varied considerably from one study to another.

Indeed, the three Big Seven studies yielded a lower degree of convergence than was evident among the Big Five studies reviewed earlier. A likely reason is less representative variable selections. As noted, Tellegen and Waller (1987) randomly selected not terms, but dictionary pages, within 25-page sections. Benet and Waller (1995) and Almagor et al. (1995) attempted to obtain a term from every fourth page. These procedures approximate representative variable selection only to the extent that the distribution of important descriptors (and their semantic content) has a rectangular (flat) distribution across pages (or parcels of pages) of a dictionary.[1] However, examination of any dictionary subverts this assumption: Some pages have many familiar person descriptors, whereas many other pages have none.[2] Also, such procedures are likely to net many unfamiliar

[1] In the same way, the distribution of U.S. Senators across American states does not approximate "representativeness" because the states do not have equal populations.

[2] I examined the same edition of *The American Heritage Dictionary* used by Tellegen and Waller (1987). Adjectives included in Goldberg's

terms (e.g., *rhadamanthine, tenebrous*), violating the frequency–importance association integral to the lexical approach. Differences between the American, Spanish, and Hebrew Big Seven studies should not be prematurely attributed to differences in culture. Instead, one must first rule out a rival hypothesis, that one or more of these studies veered far away enough from representative variable sampling to limit the potential for replication. Only the same three factors that Peabody (1987; Peabody & Goldberg, 1989) found to be the largest and most robust replicate clearly in all three of these studies; when sampling is suboptimal, larger factors are more likely than smaller factors to replicate.

The hypothesis that a wider variable selection will lead to factors beyond the Big Five is further supported by lexical studies of personality descriptors in Filipino (Tagalog). Church, Katigbak, and Reyes (1996) reported a search of a comprehensive Filipino dictionary, extracting 6,900 person-descriptive adjectives. These terms were sorted into a classification scheme expanding on that used in German lexical studies (Angleitner et al., 1990), and familiarity ratings were collected. In a further study (Church, Reyes, Katigbak, & Grimm, 1995), a set of 351 adjectives (representing traits and mental abilities) was used in a self-report task by 629 high school and college students; 9- to 12-factor solutions were originally examined. In a replication study using the same descriptors, the authors reported a replication of seven of these factors: Gregariousness (Extraversion), Concern for Others versus Egotism (Agreeableness), Conscientiousness, Self-Assurance (Emotional Stability), Intellect, Social Deviance or Negative Valence, and Temperamentalness (e.g., Irritable). Although no five-factor solution was reported, all of the Big Five were present, with Emotional Stability apparently split into two factors. Also present was a factor resembling Big Seven Negative Valence, a "name-calling" factor, drawing on social evaluation terms retained in the variable selection.

Recurring Patterns in Lexical Factor Structures

How many personality factors are there in the natural language? The results of studies in these nine languages show some recurring patterns. Three prime factors have appeared consistently in every study across nine languages, regardless of variable selection; they correspond to Extraversion, Agreeableness, and Conscientiousness (cf. De Raad & Szirmak, 1994, Table 2; Goldberg & Rosolack, 1994; Peabody & Goldberg, 1989; Saucier, 1996). Five-factor structures derived from pools of disposition descriptors include these three factors as well as Emotional Stability, and most often Intellect. When adjective pools are expanded to include a wider range of person descriptors, one sees fairly consistent evidence for an additional dimension of Negative Valence, drawing on social evaluation terms; a clear Positive Valence factor has been found in only two studies. Thus, there may be as many as seven broad orthogonal factors in these wider variable selections.

Should one prefer three, five, or seven factors? For the choice between five and seven factors, much seems to depend on how wide one prefers to extend the variable selection. For the choice between three and a larger number of factors, much may depend on one's preferred level in the hierarchical organization of attributes. Personality descriptors tend to coalesce into three large factors similar to Extraversion, Agreeableness, and Conscientiousness—the same three factors most invariant across the studies in nine languages just reviewed. The coalescence is most evident when only three factors are rotated (De Raad & Szirmak, 1994; Saucier, 1996) or when the targets of description are heterogeneous with respect to evaluation (e.g., some liked, some disliked; Peabody & Goldberg, 1989). Thus, observable effects of variable selection on factor structure seem to depend on how many factors are extracted and rotated. It follows that analyses of lexical descriptors ought to involve not just a single preferred number of factors, but a range of factor solutions, beginning with two. This procedure confers an additional advantage: When a study having less-optimal sampling is compared to other studies, convergence is more likely with the larger factors, and the potential for integrating the study with other studies is increased.

The preceding review suggests several unresolved issues: (a) whether inclusion of familiar nondisposition terms leads to additional factors, (b) whether factor structures are robust across or contingent on the breadth of variable selection, and (c) whether the multilevel array of factor structures in new, independent samples will always follow the pattern suggested by previous studies. This report addresses all three issues. I assayed the factor structure of person descriptors at various tiers from 2 to 10 factors. To enhance generalizability, I first sought a wide, representative sampling of English-language person descriptors.

Study 1: Developing an Inclusive Set of Familiar Person Descriptors

Allport and Odbert (1936) extracted 17,954 person descriptors from an unabridged dictionary and, three decades later, using a later edition, Norman (1967) added 171 terms to the Allport and Odbert set; the resulting compendium of 18,125 terms is one (exhaustive) selection of person-descriptive adjectives.[3] However, this set includes (a) a large proportion of unfamiliar terms, (b) some terms referring to race and ethnicity inappropriate for studying within-population variation, and (c)

(1982) 540-term set (and a far broader set of 500 person descriptors described in Study 2 of this article) were found disproportionately (at least three times more commonly than would be true for a rectangular distribution) on dictionary pages containing word entries beginning with the following letters: *Com-*, *Dis-*, *Ex-*, *Good-*, *Im-*, *Int-*, *Irr-*, *Pro-*, *Self-*, *Un-*, and *Well-*. It is easy to find many familiar person descriptors beginning with these letters, but they are concentrated on relatively few dictionary pages and will be systematically undersampled by page-sampling procedures. Such procedures run afoul of the disproportionate contribution of certain English prefixes to person-descriptive terms, and this problem should occur in any language that uses prefixes.

[3] A randomly selected subset of these additions: *annihilative, blue-nosed, complaisant, dedicated, growly, hard-shelled, imprecise, ingratiating, manipulative, pessimist.*

Table 1
Norman's (1967) Classification Scheme for Person Descriptors

Category	Type of terms	No. of terms	Examples
1[a] Stable traits	Prime	608	Abrupt, daring, genial, jealous, meek, persistent
2 Stable traits	Moderately difficult	566	Baleful, effusive, garrulous, obsequious
3 Stable traits	Slangy, quaint, awkward, and colloquial	2,410	Chirpy, fisty, icy, larksome, oily, ranty
4[a] Temporary states, moods, and attitudes	Prime	384	Abashed, dazed, glum, jubilant, numbed, rapt
5[a] Temporary activities	Prime	583	Babbling, enforcing, harping, jeering, obeying
6 Temporary states and temporary activities	Moderately difficult	399	Chastened, fending, jaded, nettled, opting, rankled
7 Temporary states and temporary activities	Slangy, quaint, awkward, and colloquial	1,655	Adroop, carping, fain, jawing, maddish, oathing
8[b] Social roles and relationships	Prime	242	Betrayed, employed, helpmate, leader, parental
9[b] Social effects or "social stimulus values"	Prime	163	Captivating, flustering, impressive, oppressive
10 Social roles and effects	Moderately difficult	163	Affluent, droll, galling, minion, pariah, soporific
11 Social roles and effects	Slangy, quaint, awkward, and colloquial	908	Bewearying, enthroned, heroified, kingly, pesky
12[a] Evaluative terms and mere quantifiers		760	Capable, failure, horrid, least, normal, qualified
13[a] Terms for anatomical, medical, physical, and grooming characteristics		882	Ailing, deformed, gaudy, insane, lanky, old, sedent
14 Ambiguous, vague, and tenuously or obliquely metaphorical terms		4,796	Baggage, entire, humid, marked, pathic, sprig
15 Very difficult, obscure, and little-known terms		3,606	Caltrop, floccipending, icarian, merop, peracute

[a] Indicates categories sampled in their entirety in the present studies. [b] Indicates categories sampled in entirety except for Goldberg's (1982) few deletions. Norman's categories include some nouns.

many type-nouns.[4] Given these problems, I relied on Norman's classifications of the terms, summarized in Table 1, to provide an initial basis for exclusions.

First-Round Method

I began with those 2,991 adjectives Norman (1967) had not identified as moderately or very difficult; as slangy, quaint, awkward, colloquial, ambiguous, vague; or tenuously or obliquely metaphorical; and which did not refer to race or ethnicity. I included an additional 299 terms from the further classifications of Goldberg (1982), including (a) 98 additions from a set of 424 terms referring to social roles, relationships, and effects, and (b) 201 additions from a set of 540 prime trait adjectives; from among Norman's exclusions, Goldberg had retrieved these as potentially useful descriptors.[5] Finally, I added 39 adjectives previously used in other lexical studies (Saucier, 1994c).[6] With the addition of some filler terms, 3,446 person-descriptive adjectives formed the initial pool. These adjectives were divided into five separate inventories, each having counterbalanced pages, one each for stable traits (595 terms), temporary states and activities (636 terms), social roles and effects (326 terms), evaluations (666 terms), and physical and appearance descriptors (788 terms).

Judges for the first round of familiarity ratings were recruited from class sections in upper division psychology courses at a state university on the West Coast with an ethnically diverse student population; they received extra course credit for their participation. The 83 judges included 18 men, 64 women, and 1 whose gender was not reported. Although the overall sample was preponderantly female, the most convergent raters—those whose ratings correlated most highly with the group means—were as likely to be men as women. Each subset of descriptors was rated by a subsample of between 12 and 25 judges, using a 0-to-9 (10-point) rating scale, with 0 indicating that the judge did not know the term at all, 1 indicating the term was *never*, 3 *rarely*, 5 *sometimes*, 7 *often*, and 9 *extremely often* used to describe a person.

Reliability indices were calculated, including mean interrater correlations and alpha reliabilities. In addition to the mean familiarity rating for each adjective, factor scores were calculated for each adjective on the basis of the first unrotated principal component of the intercorrelations among the judges; this index weights the response of each judge by that judge's relative degree of agreement with other judges. The mean and factor-score indices, which correlated .999 or higher, were used jointly to identify the adjectives to be retained for a second round.

[4] A randomly selected subset from the compendium of 18,125—*balky, brock, faging, inobservant, inwrapped, lightful, manged, manslaughtering, out-of-door, perhorrescing, satiny, sceneful, schnorrer, shattered, splendiferous, unsecret, unsystematic, untenty, urbane, windering*—illustrates the great preponderance of infrequently used terms.

[5] When Goldberg (1982) revised Norman's classification of these terms, he deleted 13 less useful terms: *acquainted, apparent, captive, criminal, favorite, mastered, scathing, unacquainted, unacquitted, unavoidable, unchaperoned, unreconcilable,* and *unwed.* I also deleted them.

[6] Only 2 of these 39 adjectives, *laid-back* and *soft-spoken,* actually turned out to be of sufficient familiarity to be used in Study 2.

Table 2
First-Round Familiarity Ratings: Terms, Raters, and Reliabilities

Group of terms	Raters			Reliability		
	Total	Women	Men	No. of terms	Mean r	α
Stable traits	25	17	8	843	.51	.96
Temporary states and activities	15	12	3	616	.42	.91
Social roles, relationships, and effects	12	9	3	426	.52	.92
Evaluative terms and quantifiers	18[a]	15	2	666	.59	.96
Anatomical medical, physical, and grooming terms	13	11	2	795	.60	.94
Total	83[a]	64	18	3,346		

Note. Eight adjectives were included in more than one set of terms.
[a] One rater did not indicate his or her gender.

First-Round Results

Table 2 presents reliability data for the familiarity ratings. Mean interrater correlations ranged from .42 to .60 in the five subsamples, with alpha coefficients of .91–.96.[7] Thus, there was an impressive level of interjudge agreement as to the relative familiarity of these descriptors.

The means for the 3,346 adjectives fell into a symmetrical distribution, with a mean of 4.4 and a standard deviation of 2.1. The adjectives were rank ordered by their means, and those 1,112 with a mean of 5.50 or over were retained for further study. When the adjectives were rank ordered by their factor scores, the top 1,112 adjectives included 23 adjectives that would have been excluded using the rank order by means. These 23 were also retained.

Second-Round Method

The remaining 1,135 adjectives were those with first-round means of 5.50 or over, or factor scores high enough to fall into the same rank-order range. But these means and factor scores were calculated from five different inventories: A mean of 5.50 on the trait inventory might not be precisely calibrated with a mean of 5.50 on the evaluation inventory. In the second round, the adjectives were included in a common pool.

The 1,135 adjectives from the first round were supplemented by 9 terms omitted in the first round, to form a pool of 1,144 terms. These terms were included in an 11-page inventory (104 per page) for further ratings. The 12th page of this inventory included some terms that Norman (1967) had excluded as overly ambiguous, vague, slangy, colloquial, and metaphorical. Judgments by a team of experts assisted in reducing 528 candidate terms to the 104 most promising.[8] Thus 1,248 terms were included in the second round.

Because the restricted range of relative familiarities in the second round would attenuate the interrater correlations, a larger number of judges was recruited. A sample of 46 students (28 women, 18 men) from an ethnically diverse West Coast university was supplemented by a second sample of 66 (42 women, 24 men) community residents recruited from a metropolitan area in another western state.

Each judge provided familiarity ratings, using the same 0–9 scale as in the first round, on one of two forms, randomly assigned, each containing 624 adjectives. Form 1 and Form 2 each included 6 of the 11 pages of previously rated adjectives; the page in common on the two forms enabled a check for form effects. For the community-sample judges, Form 1 was adapted so that the 104 additional adjectives were inserted in place of the page in common.

Second-Round Results

Table 3 provides reliability data for the second round. Mean interrater correlations were substantially lower than in the first round (.27–.32 instead of .42–.60) but, as a result of the larger samples, the alpha coefficients were all at least .90. For the 1,144 common terms, the ratings in the student sample had a mean of 5.5 (SD = 1.2), compared to 5.4 (SD = 1.1) in the community-resident sample. For the items rated in both rounds, almost all of which had a first-round mean of at least 5.5, the second-round means showed the expected regression toward the middle of the rating scale. For these 1,144 terms, student and community-resident means correlated .81. After the student and resident means were averaged to form an aggregate mean, the correlation with first-round ratings (for the 1,135 terms rated in both rounds) was .76. These coefficients suggest considerable agreement on a consensual rank ordering of the adjectives: Over 80% of the terms placed in the top 500 by one sample were in the top 500 in the other sample. Thus, we can expect new applications of this method to yield largely the same set of items, enhancing the likelihood of a convergent factor structure at the broad level. In contrast, page-sampling procedures tend

[7] Reliability was lowest for the temporary-states-and-activities category; this stemmed largely from lower agreement on the familiarity of temporary activity terms, which were mostly gerunds (e.g. *interfering, dancing*). It is easy to imagine lower agreement for gerunds. Some judges might have imagined a sentence stem *He/she is a ____ person* and concluded that these terms are rarely used in such a context; others might have imagined the stem *He/she is* ____ as a descriptor of behavior, and concluded that these terms are often used.

[8] The experts were all personality psychologists: Lewis R. Goldberg, Sarah E. Hampson, John M. Digman, and myself. Of the 104 terms selected by way of these expert judgments, 30 were used in Study 2: *cool, dumb, experienced, familiar, fortunate, funny, giving, handicapped, hardworking, inexperienced, innocent, lucky, mean, open, positive, powerful, private, procrastinating, rich, secure, self-conscious, sensible, short, spoiled, stuck-up, supportive, sweet, unlucky, unreasonable,* and *well-to-do.*

Table 3
Second-Round Familiarity Ratings: Terms, Raters, and Reliabilities

Group of terms	Raters			Reliability		
	Total	Women	Men	No. of terms	Mean r	α
Students						
Form 1	24	16	8	624	.28	.90
Form 2	22	12	10	624	.32	.91
Community residents						
Form 1	32	23	9	624	.32	.93
Form 2	34	19	15	624	.27	.92

Note. One hundred four terms were included on both forms for the student raters.

to produce much lower item overlap across studies and thus factor structures more reflective of idiosyncrasies of variable selection.

The set of 500 person descriptors having the highest average mean ratings can be considered a representative set of terms, given that the list of terms in each person-descriptor category is fairly comprehensive and that terms of high familiarity were included regardless of category. It comprises attributes important enough to have been encoded as single words in the language and to have a high frequency of use. However, to control systematically for breadth of variable selection, reliable classifications of these person descriptors were needed.

Classifying the Familiar Person Descriptors

Several classification schemes for person descriptors have been proposed, including the four-column system of Allport and Odbert (1936) and the 15 categories of Norman (1967); the reliabilities of these classifications were never documented. Angleitner et al. (1990) devised a synthesis of these two schemes for use in German lexical studies. The new scheme included four broad content categories: dispositions, temporary conditions, social and reputational aspects, and overt characteristics and appearance. Descriptors not fitting into any of these categories were classified as terms of limited utility, that is, either (a) context-specific and technical terms or (b) metaphorical, vague, or outmoded terms. Each of the first four broad categories included two to four subordinate categories, listed in the first column of Table 4. Angleitner et al. (1990) used this system in their classifications of 5,101 German adjectives by 10 judges. As Table 4 indicates, the first four broad categories in German generated reliability coefficients of .83 to .91 (corresponding to mean interrater correlations of .30 or more); coefficients for four of the 13 subordinate categories also were in this range.

The four broad categories were generally more reliable, and these resemble the four columns of Allport and Odbert (1936), although with a few important differences. Angleitner et al. (1990) brought abilities into the stable trait column, bodily states into the temporary conditions column, and attitudes and worldviews into the social evaluations column; having removed most of the miscellany from the fourth column, they reserved it for overt characteristics and appearance terms. This sensible scheme has been adopted in other lexical studies (Church et al., 1996; Hrebickova et al., 1995).

Classification Method

Fifteen graduate students (13 women and 2 men) were recruited from a course in psychological assessment at an ethnically diverse West Coast university and were given extra credit for participating. The task took several hours, and judges were instructed to take a break approximately once an hour.

The person descriptors that were classified included the 525 high-familiarity adjectives identified in Study 1, plus those 15 adjectives that were included in a brief Big Five inventory (Saucier, 1994a) but were not already in the 525. These 540 adjectives were printed individually on cards. Each judge received a deck of 540 cards (in a different randomized order) and a rating board—an oversized sheet of paper depicting the hierarchical system of five broad and 13 specific categories in the Angleitner et al. (1990) system. Judges were given written instructions, directing them to place each card (adjective) first into one of the five broad categories, and then into the most appropriate of the 13 subordinate categories; they also received 40 blank cards that, when filled in, would enable multiple classifications of some terms. Descriptions of the categories generally resembled those from the Angleitner et al. (1990) study; the adjectives cited as examples for each category were English translations of highly prototypical terms for that category in the German study.[9] Upon completion of their task, each judge banded together the cards in each category and returned the packets. The investigator and a colleague (Lewis R. Goldberg) each made an independent "expert" classification of the same 540 adjectives.

For each judge, an 18 × 540 matrix of classifications was constructed, enabling interrater correlations for each category. The fifteen 18 × 540 matrices were combined to form an aggregate table; in this table the prototype score for each adjective in each category was the number of judges assigning the adjective to that category (as in Angleitner et al., 1990).

The Resulting Classification

The far right column of Table 4 presents the coefficients of interjudge agreement, and the agreement between the classifications of the judges and the two experts. Interjudge agreement for the first four broad categories was .88, .93, .88, and .97 (M = .92), comparing favorably to .84, .86, .83, and .91 (M = .86) in the German study. As in the German study, the reliabilities of the subordinate categories were noticeably lower; the mean coefficient was .84 as compared to .79 for the German study. In both studies, the same three subordinate categories (behavioral states, roles and relationships, and social effects) elicited the lowest reliabilities, as well as the lowest correlations between the pooled judges' and the pooled experts' scores. Though there are weaknesses at the subordinate level, this classification scheme has great strength at the broad level.

For these reliable broad categories, illustrative highly prototypical adjectives were the following: dispositions—*creative, skillful, stubborn, hot-tempered;* temporary conditions—*anxious, happy, busy, exhausted;* social and reputational aspects

[9] Translations were obtained from Angleitner, John, and Ostendorf (n.d.); Ostendorf (1990b); or from a German–English dictionary.

Table 4
Reliability (Interjudge Agreement): Thirteen Specific and Five Broad Categories for Person-Descriptive Adjectives

Category	German	American	Correlation with experts
Dispositions	.84	.88	.70
Temperament and character traits	.78	.84	.69
Abilities, talents, etc.	.82	.91	.78
Temporary conditions	.86	.93	.81
Experimental states	.87	.93	.84
Physical–bodily states	.88	.90	.79
Behavioral states, observable activities	.63	.70	.49
Social and reputational aspects	.83	.88	.67
Social roles and relationships	.70	.68	.31
Social effects	.63	.77	.58
Pure evaluations	.82	.78	.69
Attitudes and worldviews	.90	.87	.80
Overt characteristics and appearance	.91	.97	.91
Anatomy, constitution, morphology	.90	.95	.95
Appearance, deportment, etc.	.93	.93	.77
Terms of limited utility	.75	.67	.27
Context specific or technical	.80	.66	.40
Ambiguous, vague, outmoded	.53	.56	—

Note. Reliabilities (interjudge agreement) are alpha coefficients based on intercorrelations among 15 American judges across 540 familiar English adjectives, and 10 German judges across 5,101 German adjectives (source: Angleitner, Ostendorf, & John, 1990). Correlations with experts are between the prototype scores of the 15 student judges and the independent prototype scores of two experts. The dash indicates that neither expert assigned any terms to this category.

(social evaluations)—*admirable, wonderful, interesting, respected;* overt characteristics and appearance—*skinny, tall, big, little.* Overall, Big Five marker adjectives (Saucier, 1994a) tended to be classified as dispositions, although Emotional Stability markers were more often classified as states (mean prototype score = 7.75) than as dispositions (M = 4.13), and Agreeableness markers were more often classified as social evaluations (M = 6.88) than as dispositions (M = 4.25).

Study 2: The Structure of Person Descriptors in Four Variable Selections

In Study 1, I generated a representative set of person descriptors that are both familiar and wide in their range of content, and established prototype scores for these descriptors within a reliable classification system. In Study 2, I assayed their factor structure, using varying criteria for variable selection, and rotating varying numbers of factors.

Method

Variables. The 500 most familiar descriptors from Study 1 included 14 terms of questionable appropriateness for describing the characteristics of specific people.[10] I removed these 14 terms from the stimulus set and substituted the next 14 acceptable items in rank order. The resulting 500 adjectives included 89 prototypical disposition terms, 85 prototypical state terms, 183 prototypical social-evaluations, and 38 prototypical physical–appearance terms; for the remaining 105 terms, fewer than half of the classification judges agreed on the assignment to any category.

The 500-adjective set defined the widest lexical variable selection. Definitions of personality consonant with this selection might envision either (a) a comprehensive structure of traits, both of personality and of physiognomy, or (b) a comprehensive structure of *person perception,*[11,12] including all observable person-descriptive attributes that might contribute to impressions and stereotypes. This selection of all terms was compared with three narrower selections, each representing an alternative set of reasonable boundaries for the domain of personality descriptors. Separate analyses of the terms within each descriptor category (e.g., states only or evaluations only) are beyond the scope of this article.

Though not explicitly excluding such terms, the Big Seven lexical studies reviewed earlier included few terms denoting physical and appearance characteristics. To better match their selection, in the present study a set of 455 nonphysical terms was constructed by excluding

[10] The terms were *acceptable, bald, dead, eligible, expert, female, hungry, mad, male, pregnant, qualified, sick, smiling,* and *welcome.* Each lacks either adjectivalness, applicability to anyone of either sex, or clarity without having the context specified. A list of the final 500 terms is available from me.

[11] Levy (1970) suggested the value of "morphological" factors in understanding personality. Physiognomy imposes certain conditions on an individual's existence. Reactions to these conditions—both in the self-image and adjustment of the individual, and in treatment of the individual by other persons based on his or her morphology—create part of the context for the individual's behavior. Examples include Adler's well-known view that much of human personality is an attempt to compensate for physical inferiority, and some evidence linking delinquency and physiognomic traits (Glueck & Glueck, 1956). Levy noted that Hippocrates, Rostan, and Kretschmer put forward morphological theories of temperament prior to that of Sheldon (1944).

[12] Schneider (1973), for example, noted that there is "no reason why any attribute of persons cannot be used in an implicit personality theory paradigm" (p. 305) because there is little evidence as to "what the natural units of person cognition are" (p. 306).

those 45 terms with Study 1 prototype scores of 8 or higher for overt characteristics and appearance.

Goldberg (1982, 1990, 1992) used fairly inclusive criteria for what might be considered a trait; to approximate this selection, a set of 252 broad disposition terms included those with dispositions prototype scores of 3 or higher; these are terms that can be, but are not necessarily always, thought of as referring to dispositions.[13]

Most adjectives denoting temporary conditions (e.g., *joyful, sad*) can also denote dispositions. To investigate the effects of adding state terms only to disposition terms, a fourth selection was of those 239 disposition and state terms with prototype scores on dispositions and temporary conditions that summed up to at least 8.

To provide a concurrent measure of the Big Five personality factors, adjectives were added from the Mini-Markers (Saucier, 1994a), a brief but reliable subset of Goldberg's (1992) Big Five marker scales. Because 25 of these 40 adjectives were already included in the top-ranked 500, only the 15 additional adjectives were required. Participants who provided self-ratings in the present study had also provided self-ratings (2 years earlier) on Goldberg's (1992) Big Five markers and (1 year earlier) the Revised NEO Personality Inventory (NEO-PI–R; Costa & McCrae, 1992).

To yield a total set of 525 adjectives, 10 filler adjectives were included. The 525 terms were printed on a 7-page (75 per page) form, with a 7-point response scale for each adjective. Participants were asked to indicate how "characteristic, usual, or typical" the attribute denoted by each adjective was for them.

Participants. The 525-adjective inventory, with self-report instructions, was mailed to members of a community sample consisting of residents of a medium-sized western city. Of 723 who returned the inventory, 700 (400 women and 300 men) provided protocols with 5% or fewer missing responses. The average age of these respondents was 52 years (SD = 13). In addition, the inventory was completed by 215 students at an ethnically diverse university and two community colleges, all in western states. Participants were instructed to describe someone they knew well, of the same sex and about the same age as themselves. In ratings on a 5-point scale of the degree to which the participant liked the target, only eight of the targets were placed at the two lowest points on the scale (liked *not at all* or *only a little*); these eight outlier cases were omitted. Of the remaining 207, 201 had 5% or fewer missing responses. Each participant's responses were z scored (ipsatized) to remove individual differences in use of the rating scale. Except as noted, all analyses utilized ipsatized ratings.

Analyses. Stable factors are not only more replicable, but are also less likely to be affected by alterations in method of factor extraction or rotation. The stability and robustness of a factor structure can be determined either (a) in replication studies that use different data sets or (b) within a data set by internal cross-validation—splitting the data set in half and comparing the factor structures derived from each half. Both types of methods were used here.[14] First, as an index of between-sample factor stability, principal-components solutions with varimax rotation (2–10 factors, in each of the four variable selections) based on the total self sample (N = 700) were compared to corresponding factors based on the total acquaintance sample (N = 201), by Tucker's factor congruence coefficient, which roughly resembles a correlation (Harman, 1976).[15] Second, as an index of within-sample factor stability in the larger self sample, principal-components solutions with varimax rotation of from 2 to 10 factors, for each of the four variable selections (dispositions, dispositions–states, nonphysical, and all terms) were performed on random halves of the data set. The variables were split by random selection into halves (for each of the variable selections, for each number of factors), and factor scores were generated. These two sets of factor scores were then correlated with one another. Under either index, each factor from one sample (or subsample) was matched with one from the other sample (or subsample), so as to maximize the magnitude of coefficients across the set of factors. The mean stability for a factor solution was the average of these coefficients.

For comparison purposes, I also generated stability coefficients using Everett's (1983) method, splitting the participants into random halves. Everett's method generated coefficients with a similar pattern to those from the other two methods, but they were systematically higher, impeding precise comparisons; this method is not emphasized in this article.

What are the effects of variable selection, holding the sample and the number of factors constant? I examined correlations between the factor scores generated within each of the variable selections, at each number-of-factors tier and within each full sample, matching high-correlating pairs of factors as described above. Finally, to relate the solutions to the Big Five factor structure, I examined their correlations with the varimax-rotated factor scores from a principal-components analysis of the 40 Mini-Marker adjectives and of the two other five-factor measures.

Results

Figure 1 presents the initial 25 eigenvalues within each of the four selections in the self sample. These scree plots (Cattell, 1966) show a distinct elbow after three factors (cf. Almagor et al., 1995; Benet & Waller, 1995) in all four variable selections, and another elbow after five factors in three of the selections.

To delineate any general trend in the various factor-stability coefficients, I averaged the between-sample (Tucker) and within-sample (split-half) stability coefficients into a single estimate of factor reliability for each factor solution. Table 5 depicts the between-sample and within-sample coefficients, as well as the averaged estimate.

The wider variable selections in Table 5 have a common pattern: The coefficients substantially decrease, well below .70, in average magnitude after the seven-factor solutions. The disposition selections show a corresponding decrease after only five-factor solutions. This contrast suggests that five factors better suit disposition selections, whereas seven better suit wider selections. Figure 2 depicts these trends, one line representing the averaged coefficients for the wider selections, and the other line representing the averaged coefficients for the disposition

[13] For a study of trait factor structure, Ostendorf (1990a) selected those 430 terms (from over 5,000 descriptors) that were classified by a majority of judges in either of the two disposition categories: (a) temperament and character traits and (b) abilities, talents, and their absence. Among the 500 familiar descriptors, only 89 were classified by a majority of judges as dispositions, probably too few to produce generalizable findings. However, a five-factor rotation of this small subset in the self sample did yield a Big Five structure. Correlations with corresponding factors from the Mini-Markers (Saucier, 1994a) were, respectively, .76, .75, .61, .64, and .75, for Factors I–V. The average of these coefficients, .70, is comparable to the validity coefficients obtained by Ostendorf (1990a).

[14] Parallel analysis (Horn, 1965; Zwick & Velicer, 1986) was also attempted as a method for determining the number of factors; however, this method suggests a large number of factors when the number of variables is large, estimating as many as 30 factors in the present analyses. It may be a useful estimate for the number of specific factors at a hierarchical level much lower than that of the broad orthogonal factors, a level with which the present study is not concerned.

[15] Analyses of the acquaintance sample involved more variables than participants. However, the huge variable-to-factor ratio, along with a comparison to results from a large sample, makes this characteristic relatively unproblematic (cf. Guadagnoli & Velicer, 1988).

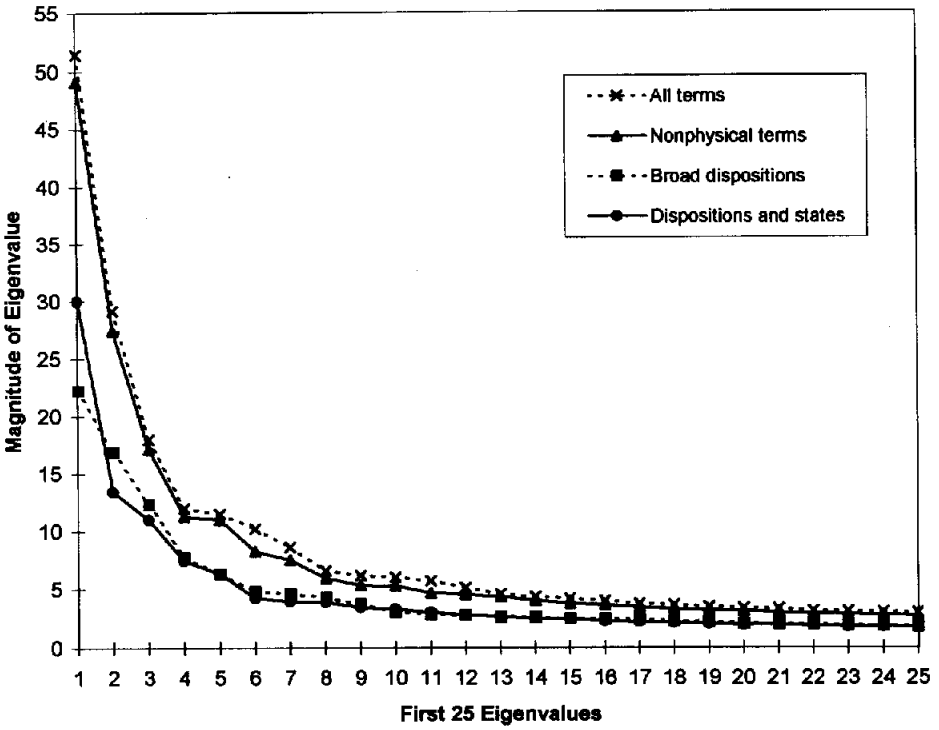

Figure 1. Plot of first 25 eigenvalues in four variable selections.

selections. There is another noteworthy pattern discernible in Figure 2. The stability estimates uniformly exceed .75 only for the three-factor solutions. Overall, stability at the two-factor level was roughly equivalent to that at the three-factor level; because factor stability is partly a function of the variable-to-factor ratio (Guadagnoli & Velicer, 1988), and because this ratio is higher (by 50%) for the two-factor solution, the stability for three factors seems more impressive than that for two.[16]

For factor solutions of 2–10 factors in the larger self sample, the effects of variable selection on the concordance between matched factors are depicted in Figure 3. This figure graphs the mean variable-selection effects, averaging across all six pairs of the four variable selections (e.g., all terms–nonphysical) within each sample.[17] The two- and three-factor solutions were strikingly invariant across variable selections, with concordance values of .95 and higher in both samples. The mean correlation between factors generated within the full set of 500 terms and the 455 nonphysical descriptors was highest: above .90 for the two-, three-, five-, six-, and seven-factor solutions, and above .80 for the remainder. This indicates that adding physical–appearance terms, so that they made up almost 10% of the variables, had relatively little effect on the factor structures.

Table 6 presents the correlations between the Big Five factors from the 40 Mini-Marker adjectives and the best matched factors from the various factor solutions in the self sample, as well as correlations between the five-factor solutions and both the 100 unipolar markers (Goldberg, 1992) and the NEO-PI–R, for each of the four variable selections.[18] When at least five factors were rotated, the Mini-Marker correlations were about .70, similar in magnitude to the Big Five replication correlations reported by Ostendorf (1990a). The first two factors corresponded to Extraversion and Agreeableness, and the third to Conscientiousness. The first five factors roughly corresponded to the Big Five. Comparatively lower correlations with the 100 Markers and NEO factors might be attributed to the time lag between assessments.

Taken together, these findings highlight the stability of three-factor solutions in all variable selections. In the two widest variable selections, the seven-factor solutions also appeared noticeably stable, whereas five factors demonstrated comparative stability in the two disposition selections. Table 7 presents, for each factor from the most stable solutions, up to 20 of those terms that had their highest loading on the factor in both samples, with each term's loading on that factor in the larger self sample.

[16] The one-factor solution, as indexed by the first unrotated principal component, was also highly stable, with .89 and higher stability coefficients across samples and across variable selections, but is not detailed in this article. The most consistent high-loading terms on this factor were (at one pole) *positive, confident, cheerful,* and *optimistic,* and (at the other pole) *insecure, negative, moody,* and *selfish.* This large factor seems to have some affinity with both constructs of Ego Resiliency (J. H. Block & Block, 1980) and General Evaluation (Saucier, 1994c).

[17] Available from me is a table presenting mean correlations between factors generated within differing variable selections in the larger self sample.

[18] For corresponding Big Five Factors I–V, Mini-Marker factors correlated .83, .59, .78, .72, and .68 with Goldberg-marker factors, and .63, .57, .63, .67, and .52 with NEO-PI–R factors. Goldberg-marker factors correlated .63, .53, .63, .70, and .53 with NEO-PI–R factors.

Table 5
Mean Stability Coefficients of Matched Factors in Four Variable Selections

Stability coefficient	Number of factors in solution								
	2	3	4	5	6	7	8	9	10
From all 500 descriptors									
Split-half variables	.94	.85	.84	.78	.73	.76	.64	.61	.58
Tucker congruence	.71	.74	.67	.58	.66	.67	.61	.62	.62
M	**.83**	**.80**	**.76**	**.68**	**.70**	**.72**	**.63**	**.62**	**.60**
From 455 nonphysical descriptors									
Split-half variables	.95	.79	.55	.77	.77	.74	.63	.60	.54
Tucker congruence	.70	.70	.73	.61	.63	.64	.62	.60	.65
M	**.83**	**.75**	**.64**	**.69**	**.70**	**.69**	**.63**	**.60**	**.60**
From 252 broad disposition descriptors									
Split-half variables	.89	.79	.71	.59	.65	.59	.63	.56	.51
Tucker congruence	.68	.75	.68	.76	.69	.61	.65	.72	.66
M	**.79**	**.77**	**.70**	**.68**	**.67**	**.60**	**.64**	**.64**	**.59**
From 239 disposition and state descriptors									
Split-half variables	.67	.77	.65	.70	.59	.58	.49	.45	.50
Tucker congruence	.71	.76	.65	.74	.70	.67	.63	.69	.68
M	**.69**	**.77**	**.65**	**.72**	**.65**	**.63**	**.56**	**.57**	**.59**

Note. Split-half variables coefficients were computed within the self sample (N = 700). Tucker coefficients were computed to compare factor loadings in the self sample to those generated in the acquaintance sample (N = 201). All coefficients not in boldface are an average of coefficients between from 2 to 10 pairs of matched factors.

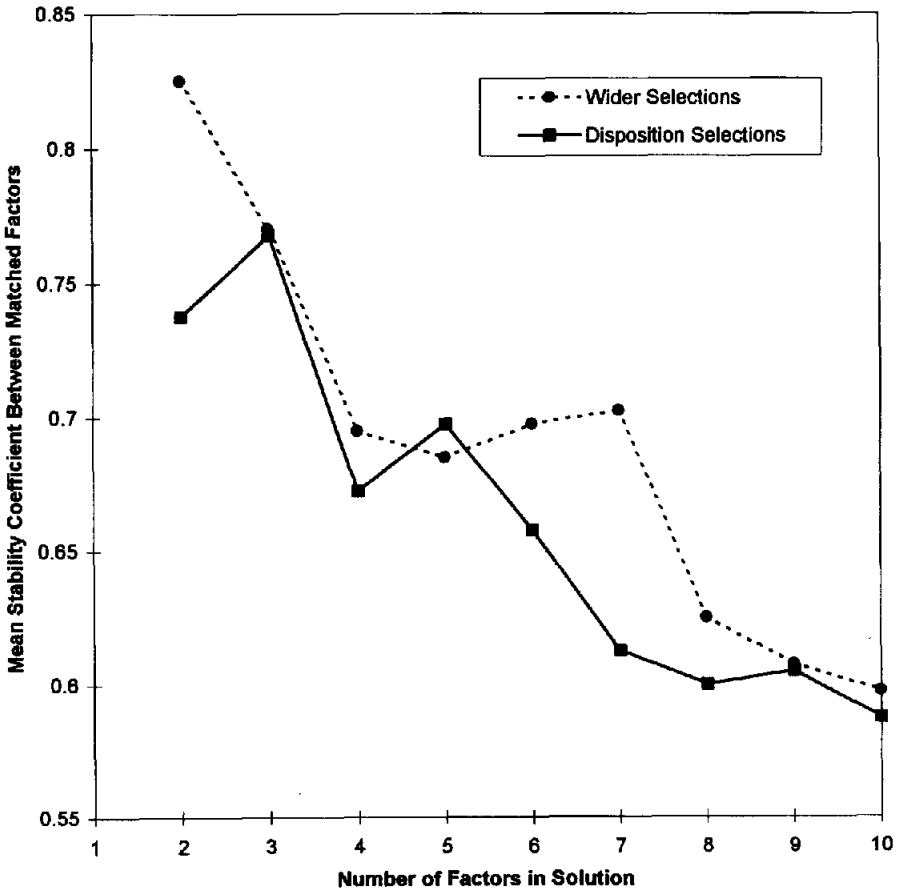

Figure 2. Estimates of factor stability for four variable selections.

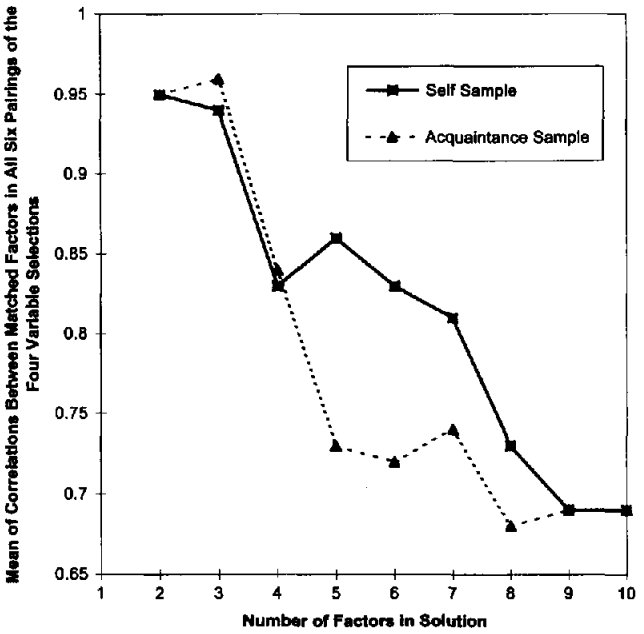

Figure 3. Variable-selection effects in two samples.

General Discussion

The three-factor solution not only demonstrated relatively high stability within and between samples (Figure 2), but was nearly identical in all four variable selections within each sample (Figure 3). The robustness of these three factors is supported by other lines of evidence. They correspond most closely to the first three of the Big Five, with content from the other two Big Five factors spread among them (cf. Saucier, 1996). And as noted earlier, studies in nine languages agree in finding at least these three factors. The studies of Peabody (1987; Peabody & Goldberg, 1989) found the same three to be the largest factors in various types of lexical data. To distinguish these three from the somewhat less broad Big Five factors, hereafter I refer to them as *mega-factors.*

Most likely, these three mega-factors, or some rotational variant of them, can be discerned in other lexical data sets; De Raad and Szirmak's (1994, Table 2) three-factor solution especially resembles that found in the present acquaintance sample. However, the three mega-factors do not correspond to Eysenck's (1970) three-factor model. One hypothesis is that Eysenck's factors are a rotational variant; however, correlations of these factors with lexical three-factor structures (De Raad & Szirmak, 1994; Saucier, 1996) do not provide good support for this hypothesis.

The two-factor solutions are best understood in the context of the three-factor solutions. In the self sample, one mega-factor, the one most related to Extraversion, was found also in the two-factor solution, whereas the other two mega-factors combined to form a single and even broader factor (see Digman, 1997, for a similar two-factor solution). By comparison, in the acquaintance sample the two factors were seen in positions that were substantially rotated within the three-dimensional space. Across samples, one factor consistently referred to degree of Social Stimulus Value (e.g., Exciting, Lively, Good-looking), whereas the other referred to degree of Respect for Social Standards (e.g., Mature, Responsible, Honest; cf. Digman, 1995). Neither of these two-factor solutions corresponded precisely to the two-factor agency-and-communion model proposed by Wiggins (Wiggins & Trapnell, 1996), based on the interpersonal circle. Instead, the axes of the interpersonal circle can be related to the first two of the mega-factors as presented in Table 7 (cf. Saucier, 1992). That the two-factor solutions tended to vary within three-space (cf. Peabody & Goldberg, 1989, Figure 1) suggests the relative primacy of the three factors.

The five-factor solutions in the disposition variable selections generated recognizable Big Five factor structures, the most stable being in the Dispositions-and-States selection. The addition of terms denoting states seems to have few consequences for the factor structure of dispositional descriptors. Unlike in many previous studies (e.g., Goldberg, 1992), a substantial number of terms loaded on the desirable pole of Emotional Stability. The version of Factor V in Table 7 is clearly an Intellect rather than an Openness factor and thus more congruent with studies in English, German, Czech, and Filipino than with studies in Dutch, Italian, or Hungarian.

The seven-factor solutions were quite similar in the two widest variable selections, and added two new factors beyond the Big Five. The sixth factor was obviously Attractiveness, comprising not only perceived physical attractiveness, but also other features that might make one attractive; such social evaluation adjectives as *glamorous, charming, graceful, seductive, delightful, fascinating,* and *terrific* all had substantial loadings (>.40) on this factor. Additional loadings over .40 for terms like *young* and *slim* further suggest that this factor has multiple facets, straddling the boundary between social evaluations and appearance descriptors. Although the factor's content does not suggest stable, internal traits, but rather the effect one has on others, perceptions of attractiveness have been the subject of much previous research attention (Eagly, Ashmore, Makhijani, & Longo, 1991; Feingold, 1992), have contributed to the content of some measures (e.g., J. Block, 1961/1978), and have been difficult to relate to the Big Five (Henss, 1996; Lanning, 1994). The present studies confirm not only their person-descriptive importance, but also their relative independence from the Big Five.

The high-loading adjectives on the seventh factor were mostly very undesirable attributes, those used in invective, denigration, vilification, and name calling (Church et al., 1995). Indeed, they included the most potentially insulting labels in the stimulus set (e.g., *insane, corrupt, evil, disgusting, stupid*). The single highest loading term across the two samples was *good-for-nothing*; had profanity been included in the stimulus set, profane descriptors would have probably be associated with this factor. However, some of the terms (e.g., *homeless, blind*) with positive loadings on this factor are not really invective. A broader commonality among the terms loading on this factor—applying to both *homeless* and *evil*—is that they represent low base-rate attributes. Indeed, for the insults involved to have any force, invective must denote low base-rate attributes, although not all low base-rate attributes must be invective. Statistically, all low base-rate attributes will tend to produce a skewed response dis-

Table 6
Correlations of Big Five Mini-Marker Factors With Corresponding Lexical Factors

Lexical factor solution	Factors from Big Five markers I	II	III	IV	V	*M*
From all 500 descriptors						
Two-factor	.55	.61	—	—	—	.58
Three-factor	.58	.77	.49	—	—	.61
Four-factor	.51	.73	.55	.64	—	.61
Five-factor	.60	.78	.63	.69	.66	.67
Six-factor	.71	.82	.63	.71	.67	.71
Seven-factor	.76	.80	.68	.67	.68	.72
Eight-factor	.76	.80	.68	.67	.66	.71
M	.64	.76	.61	.68	.67	
From 455 nonphysical descriptors						
Two-factor	.56	.61	—	—	—	.58
Three-factor	.53	.77	.43	—	—	.58
Four-factor	.51	.84	—	.60	.59	.64
Five-factor	.46	.80	.53	.60	.67	.61
Six-factor	.73	.78	.58	.68	.67	.69
Seven-factor	.75	.80	.69	.67	.66	.71
Eight-factor	.73	.79	.67	.68	.64	.70
M	.61	.77	.58	.65	.64	
From 252 broad disposition descriptors						
Two-factor	.54	.66	—	—	—	.60
Three-factor	.60	.78	.48	—	—	.62
Four-factor	.72	.77	.58	—	.62	.67
Five-factor	.67	.74	.60	.68	.71	.68
Six-factor	.72	.74	.66	.69	.71	.70
Seven-factor	.74	.77	.73	.67	.73	.73
Eight-factor	.81	.80	.73	.68	.71	.75
M	.69	.75	.63	.68	.70	
From 239 disposition and state descriptors						
Two-factor	.55	.61	—	—	—	.58
Three-factor	.58	.77	.49	—	—	.61
Four-factor	.51	.73	.55	.64	—	.61
Five-factor	.60	.78	.63	.69	.66	.67
Six-factor	.71	.82	.63	.71	.67	.71
Seven-factor	.76	.80	.68	.67	.68	.72
Eight-factor	.76	.80	.68	.67	.66	.71
M	.64	.76	.61	.68	.67	
Correlations of other five-factor measures with the five-factor solutions						
100 unipolar markers (Goldberg, 1992) administered 2 years prior (N = 694)						
All 500 terms	.55	.59	.59	.62	.62	.59
Nonphysical terms	.42	.61	.51	.52	.62	.54
Broad dispositions	.70	.55	.55	.62	.61	.61
Dispositions and states	.70	.61	.55	.66	.62	.63
NEO-PI–R Domain Scales (Costa & McCrae, 1992) administered 1 year prior (N = 556)						
All 500 terms	.63	.75	.44	.74	.40	.59
Nonphysical terms	.43	.75	.35	.70	.42	.53
Broad dispositions	.38	.54	.42	.46	.41	.44
Dispositions and states	.53	.65	.41	.71	.41	.54

Note. N = 700, except as noted. Big Five marker factors were derived from 40 adjectives (Saucier, 1994a). Dashes indicate that no lexical factor in the solution had its highest correlation with that Big Five factor. NEO-PI–R = Revised NEO Personality Inventory.

tribution; the 20 terms with the most extreme skew in the self sample all had positive loadings of .30 and over on this factor (Waller, in press, noted a similar tendency).

This seventh factor resembles the Negative Valence factor identified by Benet and Waller (1995) and Almagor et al. (1995). But Negative Valence seems an imprecise label: Many adjectives with virtually zero loadings on this factor (e.g., *depressed, negative, stingy, careless, snobbish, clumsy*) also have negative valence and, as some of the analyses of Almagor et al. indicate, there is sometimes a positive pole to this factor.

This seventh factor showed some instability across procedures. In seven-factor rotations using raw (nonipsatized) data in the self sample, the factor was absent, whereas the first six factors remained virtually the same; it did, however, replicate

Table 7
Best Replicating Adjectives for Factors in the Most Stable Solutions in Study 2

Solution	I^B	II^B	III^B
Three factors (broad dispositions selection)	.62 Self-confident	.67 Kind-hearted	.58 Practical
	.61 Confident	.63 Kind	.57 Sensible
	.49 Assertive	.62 Warm-hearted	.53 Realistic
	.47 Persuasive	.61 Warm	.52 Dependable
	.45 Bold	.58 Giving	.52 Responsible
	.37 Persistent	.58 Cheerful	.51 Reliable
	.37 Strong	.57 Friendly	.50 Careful
	.37 Competitive	.56 Gentle	.50 Trustworthy
	.37 Energetic	.56 Caring	.49 Honest
	.36 Brave	.54 Pleasant	.45 Logical
	.35 Forward	−.42 Unfriendly	.45 Prompt
	.34 Direct	−.42 Obsessive	.44 Conscientious
	.32 Straightforward	−.43 Hot-tempered	.44 Cautious
	.31 Outspoken	−.45 Negative	.43 Ethical
	.31 Adventurous	−.45 Argumentative	.42 Stable
	.30 Tough	−.46 Demanding	.42 Truthful
	−.39 Self-conscious	−.46 Critical	.40 Serious
	−.41 Shy	−.49 Arrogant	−.29 Unpredictable
	−.43 Bashful	−.50 Self-centered	−.33 Childish
		−.54 Selfish	−.35 Reckless

	I: Extraversion	II: Agreeableness	III: Conscientiousness	IV: Emotional Stability	V: Intellect
Five factors (dispositions and states selection)	.52 Outgoing	.66 Kind-hearted	.50 Organized	.65 Relaxed	.69 Smart
	.49 Assertive	.66 Warm-hearted	.49 Prompt	.64 Happy	.69 Intelligent
	.48 Aggressive	.64 Caring	.48 Thorough	.62 Secure	.63 Intellectual
	.43 Talkative	.62 Compassionate	.48 Consistent	.61 Satisfied	.63 Knowledgeable
	.42 Bold	.61 Giving	.48 Efficient	.58 Carefree	.59 Talented
	.41 Energetic	.60 Warm	.46 Neat	.56 Peaceful	.57 Brilliant
	.40 Excited	.60 Kind	.43 Punctual	.54 Optimistic	.54 Bright
	.37 Outspoken	.57 Sympathetic	.41 Alert	.54 Easygoing	.53 Wise
	.35 Competitive	.54 Gentle	.27 Awake	.50 Self-confident	.52 Gifted
	.30 Uninhibited	.47 Sensitive	−.28 Irrational	.50 Comfortable	.50 Skilled
	.28 Direct	.46 Affectionate	−.37 Noisy	−.54 Annoyed	.49 Thinking
	.24 Brave	.44 Generous	−.38 Childish	−.54 Moody	.49 Imaginative
	−.34 Alone	.44 Thoughtful		−.57 Irritated	.48 Skillful
	−.37 Embarrassed	.40 Romantic		−.57 Tense	.47 Clever
	−.39 Sleepy	.38 Sentimental		−.59 Depressed	.47 Educated
	−.46 Quiet	.36 Forgiving		−.59 Disappointed	.46 Creative
	−.52 Bashful	−.35 Critical		−.59 Upset	−.26 Inexperienced
	−.54 Shy	−.38 Mean		−.59 Frustrated	−.30 Naive
		−.43 Arrogant		−.65 Sad	−.31 Ignorant
		−.44 Selfish		−.65 Troubled	−.40 Uneducated

	I: Extraversion	II: Agreeableness	III: Conscient.	IV: Emotional Stability	V: Intellect	Sixth factor	Seventh factor
Seven factors	.57 Outgoing	.66 Kind-hearted	.55 Organized	.63 Happy	.59 Intelligent	.65 Beautiful	.57 Evil
	.54 Talkative	.63 Warm-hearted	.54 Neat	.59 Relaxed	.56 Intellectual	.64 Good-looking	.56 Good-for-nothing
	.47 Outspoken	.62 Kind	.48 Prompt	.55 Optimistic	.55 Brilliant		.53 Insane
	.45 Forward	.60 Caring	.43 Strict	.55 Carefree	.55 Knowledgeable	.63 Gorgeous	.52 Terrible
	.43 Expressive	.59 Compassionate	.43 Proper	.52 Positive	.55 Exceptional	.62 Attractive	.52 Corrupt
	.43 Loud	.58 Giving	.42 Punctual	.52 Joyful	.54 Talented	.59 Cute	.51 Awful
	.42 Bold	.47 Considerate	.37 Useful	.51 Cheerful	.52 Excellent	.57 Glamorous	.50 Cruel
	.40 Noisy	.47 Motherly	−.26 Abnormal	.50 Pleased	.52 Gifted	.57 Pretty	.49 Retarded
	.39 Dominant	.46 Supportive	−.38 Strange	.48 Easygoing	.50 Extraordinary	.56 Sexy	.47 Dumb
	.37 Colorful	.45 Helpful	−.40 Sloppy	.47 Comfortable	.49 Skilled	.55 Appealing	.47 Bad
	.36 Excited	−.30 Materialistic	−.42 Messy	−.43 Grumpy	.48 Wise	.55 Lovely	.46 Disgusting
	.32 Uninhibited	−.32 Conceited	−.47 Weird	−.44 Crabby	.45 Skillful	.53 Adorable	.45 Pathetic
	.28 Direct	−.35 Snobbish		−.47 Bitter	.45 Outstanding	.52 Desirable	.44 Dangerous
	−.30 Private	−.35 Inconsiderate		−.49 Annoyed	.45 Thinking	.47 Terrific	.44 Homeless
	−.30 Embarrassed	−.37 Greedy		−.50 Negative	.45 Remarkable	.47 Seductive	.43 Abusive
	−.52 Bashful	−.40 Cocky		−.51 Moody	−.27 Weak	.45 Delightful	.41 Senile
	−.54 Soft-spoken	−.44 Stuck-up		−.52 Irritated	−.29 Inexperienced	.45 Great	.40 Incompetent
	−.56 Shy	−.46 Self-centered		−.55 Tense	−.31 Naive	−.28 Big	.38 Violent
	−.59 Quiet	−.50 Selfish		−.57 Angry	−.33 Narrow-minded	−.38 Chubby	.37 Frightening
		−.50 Arrogant		−.57 Upset	−.38 Dependent	−.45 Fat	.36 Blind
						−.53 Unattractive	.36 Stupid

Note. Loadings are from the self sample (N = 700). Adjectives are those with highest loading on factor in both samples. I^B, II^B, and III^B refer to the mega-factor counterparts of Extraversion, Agreeableness, and Conscientiousness (conscient.), respectively.

in the raw acquaintance-sample data.[19] Further studies are required to assess the robustness of this factor in nonipsatized data. In developing scales to measure this factor, extreme evaluative items cannot be avoided, as items that are either less extreme or more clear in descriptive reference are likely to load on one of the other content factors. For example, Saucier (1994b; cf. McCrae & Costa, 1995) reported that one adjective measure of Negative Valence (Waller & Zavala, 1993) was confounded with (low) Big Five Agreeableness.

Broader Implications of These Studies

The findings tend to confirm the recurrent patterns evident in previous lexical factor studies, but bring these patterns into clearer focus. There may be no single superior "magic number" of factors. Depending on variable-selection preferences, there are reasonable arguments for the "magicality" of the numbers three, five, or seven; Goldberg's (1983) thesis that there are five personality factors, "plus or minus two" may need restating as "five plus or minus exactly two."

Instead of asking "what is the correct number of factors?" it may be more useful to ask a different question: As one sequentially increases the number of factors, how does the pattern of factor emergence in one study correspond to that in other studies? My findings suggest an integrative multitiered hierarchical representation of orthogonal factors, which seems to correspond to the recurring pattern of factor emergence in previous studies. On the basis of lexical evidence, Extraversion, Agreeableness, and Conscientiousness are the prime factors of the Big Five: They generalize across languages and anchor an impressively stable structure of three broader factors. But some trade-offs are involved: Although the few broader factors may be more stable and impervious to variable-selection effects, fewer factors means less information. At a slightly more differentiated and information-rich level, the Big Five structure is robust, providing that disposition and state terms are emphasized, excluding most social evaluation and appearance descriptors. On the other hand, if one elects to include many such descriptors, a seven-factor structure appears to be more stable. Of these seven, five factors (the Big Five) are composed mainly of disposition descriptors, and two factors mainly of social effect (e.g., *attractive*) and social evaluation (e.g., *evil*) terms.

However, the seven-factor structure obtained in Study 2 does not correspond in two respects to the Big Seven: It contained (a) an Intellect rather than a Conventionality factor and (b) an Attractiveness factor instead of Positive Valence. The 500 terms used in Study 2 included 14 previously identified with Positive Valence (Benet & Waller, 1995; Waller, in press; Waller & Zavala, 1993). In the larger self sample, nine of these terms had their highest loadings on the Intellect factor, these loadings being .43 and higher (*excellent, exceptional, gifted, outstanding, skilled, talented, remarkable* on the desirable pole, and *average* and *ordinary*, as expected, on the undesirable pole). The other five terms (*admirable, impressive, influential, important,* and *powerful*) had their highest loadings, .37 and higher, on the Attractiveness factor. In general, the Positive Valence terms were blends of high Intellect and high Attractiveness. Some Big Seven studies (Benet & Waller, 1995; Tellegen & Waller, 1987) apparently included too few Intellect and Attractiveness terms to detect this pattern; the Positive Valence factor of Almagor et al. (1995) can be considered an Intellect factor. Given these substantial associations with both Intellect and Attractiveness, if either Intellect or Attractiveness is well represented in a variable selection, one can expect that there will be no distinct Positive Valence factor.

The present studies demonstrate the usefulness of (a) including a full range of lexical person descriptors in the initial pool (e.g., Tellegen & Waller, 1987), (b) reducing a large variable pool by a familiarity criterion (e.g., Szirmak & De Raad, 1994), (c) indexing factor stability and emphasizing stable factors (e.g., Ostendorf, 1990a), and (d) assaying an array rather than just one tier of the factor structure (e.g., De Raad & Szirmak, 1994). Future lexical studies might incorporate all of these features. Helpful from an integrative standpoint would be reanalyses of data sets from previous lexical studies to assess their factor stability in factor structures of up to at least seven factors. Rather than engaging in a parochial defense of certain "magic numbers," researchers might look at the dimensions of phenotypic individual variation less simplistically, from more than one point of view. However, not all points of view are worth taking; for finding the more worthwhile points of view, factor stability is a useful guide.

[19] Instead of an Invective factor, a Conservatism factor (split off from Conscientiousness) emerged in the nonphysical selection, and a Relaxation factor (split off from Emotional Stability) emerged in the all-terms selection.

References

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

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*(1, Whole No. 211).

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.

The American heritage dictionary (2nd college ed.). (1982). Boston: Houghton-Mifflin.

Angleitner, A., John, O. P., & Ostendorf, F. (n.d.). *A taxonomy of German personality-descriptive terms.* Unpublished manuscript, Universitaet Bielefeld.

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

Benet, V., & Waller, N. G. (1995). *A Spanish taxonomy of trait terms.* Unpublished manuscript, University of California, Davis.

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

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

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 Symposium on Child Psychology* (Vol. 13, pp. 39–101). Hillsdale, NJ: Erlbaum.

Brokken, F. B. (1978). *The language of personality.* Meppel, Netherlands: Krips.

Caprara, G. V., & Perugini, M. (1994). Personality described by adjec-

tives: Generalizability of the Big Five to the Italian lexical context. *European Journal of Personality, 8,* 357–369.

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. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Church, A. T., Katigbak, M. S., & Reyes, J. A. (1996). Toward a taxonomy of trait adjectives in Filipino: Comparing personality lexicons across cultures. *European Journal of Personality, 10,* 3–24.

Church, A. T., Reyes, J. A. S., Katigbak, M. S., & Grimm, S. D. (1995, August). *Indigenous Filipino personality dimensions and the Big Five model.* Paper presented at the 103rd Annual Convention of the American Psychological Association, New York.

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

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

De Raad, B., & Szirmak, Z. (1994). The search for the Big Five in a non-Indo-European language: The Hungarian trait structure and its relationship to the EPQ and the PTS. *European Review of Applied Psychology, 44,* 17–26.

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

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

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but . . . : A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110,* 109–128.

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

Eysenck, H. J. (1970). *The structure of human personality* (3rd ed.). London: Methuen.

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

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111,* 304–341.

Glueck, S., & Glueck, E. (1956). *Physique and delinquency.* New York: Harper.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. W. 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. (1983, June). *The magical number five, plus or minus two: Some considerations on the dimensionality of personality descriptors.* Paper presented at a research seminar, Gerontology Research Center, National Institute on Aging/National Institutes of Health, Baltimore.

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., & 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* (pp. 7–35). Hillsdale, NJ: Erlbaum.

Goldberg, L. R., & Saucier, G. (1995). So what do you propose we use instead—A reply to Block. *Psychological Bulletin, 177,* 221–225.

Greenwood, J. D. (1991). *Relations and representations.* London: Routledge.

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

Harman, H. H. (1976). *Modern factor analysis* (3rd ed., rev.). Chicago: University of Chicago Press.

Henss, R. (1996, July). *The Big Five and physical attractiveness.* Paper presented at the 8th European Conference on Personality, Ghent, Belgium.

Hofstee, W. K. B., Brokken, F. B., & Land, H. (1981). Constructie van een standaardpersoonlijkheids-eigenschappenlijst (SPEL) [Construction of a standard list of personality descriptive adjectives]. *Nederlands Tijdschrift voor de Psychologie, 34,* 443–452.

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

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185.

Hrebickova, M., Ostendorf, F., & Angleitner, A. (1995, July). *Basic dimensions of personality description in the Czech language.* Poster presented at the 7th Meeting of the International Society for Study of Individual Differences, July 14–19, Warsaw, Poland.

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.

Lanning, K. (1994). Dimensionality of observer ratings on the California Adult Q-set. *Journal of Personality and Social Psychology, 67,* 151–160.

Levy, L. H. (1970). *Conceptions of personality: Theories and research.* New York: Random House.

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. (1995). Positive and negative valence within the five-factor model. *Journal of Research in Personality, 29,* 443–460.

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

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

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

Ostendorf, F. (1990b). *Taxonomie einer Liste von 830 personbeschreibenden Adjektiven* [Taxonony of a list of 830 person-descriptive adjectives]. Unpublished manuscript, Universitaet Bielefeld.

Ostendorf, F., & Angleitner, A. (1993, July). *A German replication study of the five-factor model based on a comprehensive taxonomy of personality descriptive adjectives.* Paper presented at the 6th meet-

ing of the International Society for the Study of Individual Differences, Baltimore.

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

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

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

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, August). *Positive and negative valence: Beyond the Big Five?* Paper presented at the 102nd Annual Convention of the American Psychological Association, Los Angeles, CA.

Saucier, G. (1994c). Separating description and evaluation in the structure of personality attributes. *Journal of Personality and Social Psychology, 66,* 141–154.

Saucier, G. (1996). *Personality in 3-D: Integrating three- and five-factor models.* Unpublished manuscript, California State University, San Bernardino.

Saucier, G., & Goldberg, L. R. (1996a). Evidence for the Big Five factors in analyses of personality adjectives in English. *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.

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin, 79,* 294–309.

Sheldon, W. H. (1944). Constitutional factors in personality. In J. M. Hunt (Ed.), *Personality and the behavior disorders* (pp. 526–549). New York: Ronald.

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

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

Tellegen, A., Grove, W. M., & Waller, N. G. (1991). *Inventory of Personal Characteristics #7 (IPC7).* Unpublished manuscript, University of Minnesota.

Tellegen, A., & Waller, N. G. (1987, August–September). *Re-examining basic dimensions of natural language trait descriptors.* Paper presented at the 95th Annual Convention of the American Psychological Association, New York.

Velicer, W. F., & Fava, J. L. (1987). An evaluation of the effects of variable sampling on component, image, and factor analysis. *Multivariate Behavioral Research, 22,* 193–209.

Waller, N. G. (in press). Evaluating the structure of personality. In C. R. Cloninger (Ed.), *Personality and psychopathology.* Washington, DC: American Psychiatric Press.

Waller, N. G., & Zavala, J. D. (1993). Evaluating the Big Five. *Psychological Inquiry, 4,* 131–134.

Wierzbicka, A. (1986). What's in a noun? (Or: How do nouns differ in meaning from adjectives?). *Studies in Language, 10,* 353–389.

Wiggins, J. S., & Trapnell, P. (1996). A dyadic-interactional perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 88–162). New York: Guilford.

Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research, 16,* 263–289.

Zwick, W. R., & Velicer, W. F. (1986). A comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.

Received February 19, 1996
Revision received December 9, 1996
Accepted December 11, 1996 ■