

The lexical approach to personality: a historical review of trait taxonomic research

OLIVER P. JOHN

University of California at Berkeley, U.S.A.

ALOIS ANGLEITNER and FRITZ OSTENDORF

Universität Bielefeld, Federal Republic of Germany

Abstract

We review research aimed at the development of a compelling taxonomy of personality-descriptive terms. We identify five issues central to the construction of personality taxonomies and discuss the advantages and limitations of the lexical approach. Our review of research stimulated by this approach begins with Allport and Odbert's trait names, retraces the procedures that led to Cattell's personality factors, and summarizes contemporary work in English and in Dutch. Taxonomers and lay people alike view stable traits as the most fundamental personality concepts. At the broadest level of abstraction, this domain is often represented by five dimensions related to Power (or Surgency), Love (Agreeableness), Work (Conscientiousness), Affect (Emotional Stability), and Intellect (Culture). However, given that their number and interpretation are still debated, these dimensions should not be considered final but as a heuristically useful framework. Indeed, if the fifth dimension represents Intellect, our review suggests that Culture, Values and even Autonomy-Conformity might be additional dimensions. Some researchers, such as Goldberg and Wiggins, have constructed more narrow categories to permit a differentiated mapping of trait descriptors. Taxonomies of states, activities, and social roles and effects still need to be developed; a broad conception of personality might also include health, fitness, and physical attractiveness. A comprehensive taxonomy would provide a common framework for research guided by different theoretical orientations and could guide the selection of variables for research. Ultimately, the value of a taxonomy depends on its success in predicting important outcomes in people's lives.

Correspondence concerning this article may be addressed to Oliver P. John, Department of Psychology, University of California, Berkeley, CA 94720, USA, or to Alois Angleitner, Abteilung Psychologie, Universität Bielefeld. Postfach 8640, D-4800 Bielefeld 1, Federal Republic of Germany.

A taxonomy is a systematic framework for distinguishing, ordering, and naming types and groups within a subject field. For example, in biology, the orderly classification of plants and animals evolved during the Linnean Period and firmly established the standard binomial nomenclature. First based on the static concept of fixity of the species, this taxonomy evolved considerably over the centuries and is now based on dynamic concepts, such as natural selection and other mechanisms underlying speciation. The availability of a taxonomy that is generally accepted in the field has been a tremendous asset for biologists. It permits researchers to study specified classes of instances, instead of examining separately every individual instance, and it serves to facilitate the communication and accumulation of empirical findings about these classes and their instances by offering a standard nomenclature.

Personality psychology has not yet established a generally accepted taxonomy of its subject matter, which includes all variation in the overt social behaviour and the internal experience of individuals. Thus, the task faced by the personality psychologist would seem more difficult than that faced by the early biologists. Biologists classify individual exemplars according to their attributes; in personality taxonomies, the exemplars are the attributes themselves. Whereas exemplars of animals have a discrete physical existence and can thus be 'found' by the researcher in the field, personality attributes can neither be seen nor found nor otherwise observed directly. Personality attributes are abstract concepts that have to be inferred, and even the existence of personality attributes in people has been debated by some theorists. There is another important difference between personality attributes and natural objects. Although the number of distinct animals, for example, is exceedingly large, it is finite and potentially countable at any given time. In contrast, the number of ways in which people differ from each other is seemingly infinite, as one can always think of new attributes (e.g. *computer literacy*, *nerd*) or subdivide previously known attributes into more specific ones. Given these particular difficulties, it may seem less surprising that personality psychologists continue to struggle with the task of cataloguing, ordering, and naming in a standard fashion the domain of individual differences.

BASIC ISSUES IN THE CONSTRUCTION OF PERSONALITY TAXONOMIES

In an analysis of the major taxonomic systems, John, Goldberg, and Angleitner (1984) identified five major issues in the construction of personality taxonomies, which are summarized in Figure 1. First, the taxonomers must specify the domain of phenomena to be covered by the taxonomy. This task is inherently theoretical, tantamount to defining what one means by personality. A quick perusal of the introductory chapters of personality textbooks shows little evidence of agreement on the definition of *personality*. However, in contrast to textbook authors who may simply evade the task, taxonomers have to make practical decisions concerning their definition of personality. They have to specify what should be classified before they can begin classifying it. Nevertheless, some taxonomers have specified the instances they considered relevant without explicating their rationale, that is their definition of personality.

In addition to domain specification and instance identification, the taxonomer has to decide on the scope, or inclusiveness, of the classification. Should the biologist seek one comprehensive taxonomic structure that covers all animals and

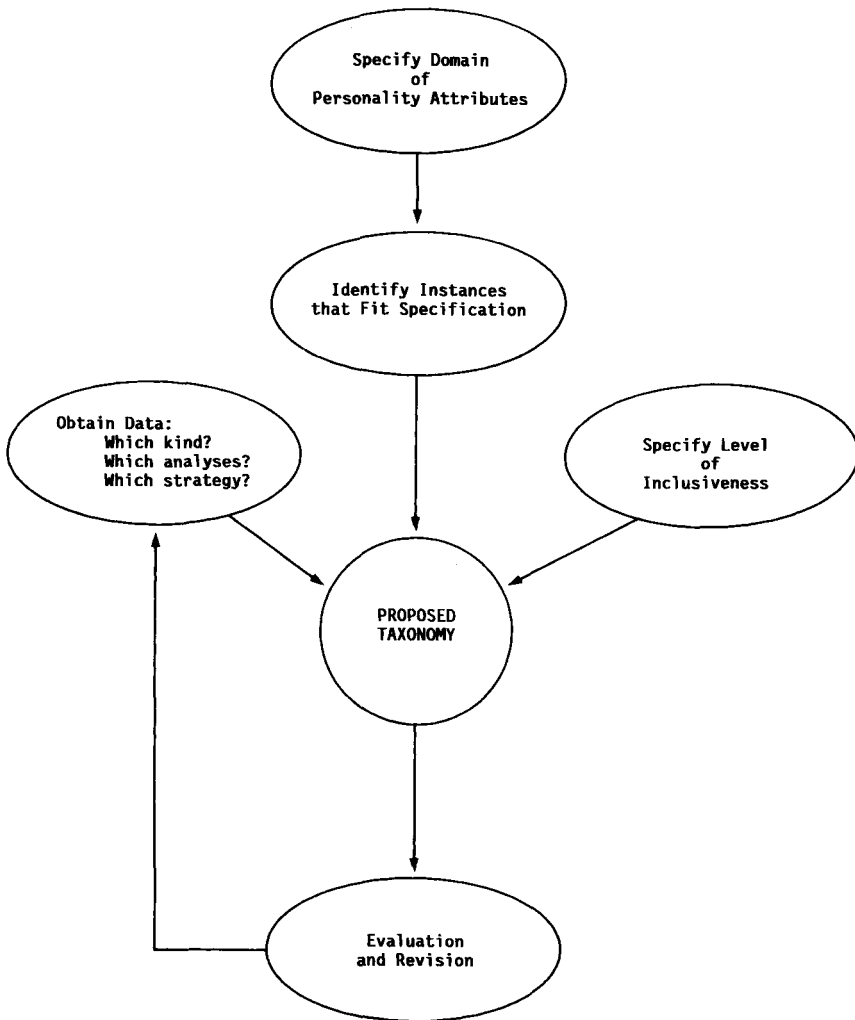


Figure 1. Five basic issues in the construction of personality taxonomies

plants? Or, would a taxonomy covering only a restricted domain be more useful? In the taxonomies described below, most researchers have restricted themselves to a limited domain (i.e. stable traits) but found themselves still faced with an overwhelmingly large number of personality attributes. Therefore, many have focused on even more narrow subsets (e.g. interpersonal traits) that they regarded as most important.

Further decisions have to be made about the procedures used to discover the structure (for those who believe that there is one 'out there' to be discovered) or construct it (for those whose world view is a social-constructivist one). Whatever one's metaphysical orientation, decisions have to be made about (a) the kind of data relevant to taxonomy construction, (b) data-analytic procedures (e.g. intuitive vs objective), and (c) strategies of construction (e.g. inductive vs deductive). The final issue involves evaluation. How should a particular taxonomic proposal be

evaluated and compared to alternative ones? How can changes be brought about once a taxonomy has been in place for some time? To be useful, the system must be reasonably stable yet sufficiently flexible so that it can accommodate new findings and ideas. Generalizability, comprehensiveness (or inclusiveness), and external validity are potential evaluation criteria of taxonomies, although so far they have been seldom used.

The second issue in Figure 1 has elicited much disagreement among personality researchers. As we noted earlier, personality attributes have to be inferred and measured. Even if one has succeeded in specifying the domain to be taxonomized, one still has to assemble the universe of instances that fit that specification. Several researchers have used an inductive approach, that is they identified attributes on the basis of individuals' responses to large sets of personality-descriptive sentences, such as questionnaire items. However, the specification of the universe of such descriptive sentences has not been very systematic (Angleitner, John and Loehr, 1986). Another route to a comprehensive set is to use clinical expertise and rely on experts to supply statements they have found useful for personality description (Block, 1961); the resulting set can then be used to describe a representative sample of individuals. Analyses of these descriptions may help identify concepts in the set that are redundant and can be eliminated, as well as others that are missing and need to be added.

Other researchers have introduced personality attributes on the basis of theoretical arguments (e.g. Snyder, 1974) or to account for experimental findings (e.g. Fenigstein, Scheier and Buss, 1975). These research efforts have highlighted instances of personality attributes that an eventual taxonomy should include or, at least, be able to include. However, a survey limited to the attributes psychologists have studied is unlikely to yield anything but a motley collection, and it would be difficult to argue that such a collection would constitute a universe of content that is well-defined or even comprehensive.¹ Even approaches that define personality attributes as categories of acts (Buss and Craik, 1983) have to identify an initial *source* from which to sample categories. One such source is the natural language, which provides a finite yet extensive set of terms that refer to differences among people. Even more conveniently, lexicographers have been cataloguing this storehouse of folk knowledge for many years.

THE RATIONALE OF THE LEXICAL APPROACH AND ITS LIMITATIONS

What sets apart the natural language as a source of personality attributes from other sources is embodied in a basic assumption that underlies the various lexical approaches. This assumption was first articulated by Klages (1926/1932), and then elaborated by Allport (1937), Cattell (1943b), Norman (1963), and Goldberg (1982): *Those individual differences that are most salient and socially relevant in people's lives will eventually become encoded into their language; the more important such a difference, the more likely is it to become expressed as a single word.* The analysis of the personality vocabulary represented in a natural language

¹Indeed, one of Guilford's contributions to the personality assessment literature was his effort to collect and integrate empirically the various measures of personality proposed over the years. However, the comprehensive factor-analytical structures he proposed (e.g. Guilford and Zimmerman, 1956) did not achieve the status of a generally accepted taxonomic superstructure.

should thus yield a finite set of attributes that the people in the language community have generally found to be the most important.

Although this rationale for the lexical approach is reasonably clear, it has some important limitations. For one, the criteria of importance that have shaped the personality lexicons of lay people are not well-understood. Indeed, such terms probably serve not only descriptive but also expressive and evaluative purposes which are not necessarily appropriate for the scientific description of personality. Moreover, some individual differences that would be of scientific interest may not be sufficiently obvious for lay people to notice and encode as single words. One way to evaluate the comprehensiveness of language-based taxonomies, therefore, is to compare them to sets of attributes compiled from the personality literature. Several taxonomers (particularly Cattell) have assigned *more* weight in determining the universe of content to scientific interest than to the social-cultural standards embodied in language.

A second limitation of the lexical rationale is that the personality attributes included in the lexicon may differ across language communities, and tend to change over time. These characteristics of natural languages may limit the generalizability of language-based taxonomies and serve as a reminder that personality attributes are difficult to study outside the cultural context in which they have emerged. Nevertheless, the effects of culture and time on the personality lexicons can be studied empirically; researchers have now begun cross-cultural studies to evaluate the generalizability of their taxonomies (John *et al.*, 1984).

A third, frequently voiced, objection to the lexical rationale is that the meaning of natural-language terms is too vaguely defined, ambiguous, and context-dependent to serve as useful scientific terminology (e.g. Bromley, 1977). The problem of fuzzy meanings has plagued personality research for many decades. Allport, for example, argued that

trait-names are range-names. That is so to say, when we designate Tom and Ted both as *aggressive*, we do not mean that their aggressiveness is identical in kind . . . words are nothing more than ratbag categories. We put into them a vast array of meanings that are roughly, but only roughly, comparable with another.

But dissatisfied (with the natural language) as we may be, verbal bags are all we have to work with. Perhaps some day psychological science will have other symbols to designate traits (factorial or mathematical), but progress in this direction to date is too slight to tempt us to abandon primary allegiance to our mother tongue. Our fate is to analyze traits in words, doing the best we can along the way to *define* our terms (Allport, 1961, pp. 355–356).

Similar philosophical views of personality attributes have been articulated by Hampshire (1953) and later Wiggins (1974), who conceptualized traits as categorical summary statements about a person's behaviour. More recently, Allport's (1937, 1961) view of personality terms as 'ratbag categories' or labels for a range of topographically dissimilar behaviours have been formalized (e.g. Buss and Craik, 1980, 1981, 1983; Hampson, 1982). Personality attributes are seen as natural categories, with general properties that are similar to other abstract categories (e.g. *red*, *chair*, *truth*). The instances of these categories are behavioural acts or events that vary in their prototypicality for a particular trait; there are no clear boundaries between adjacent categories. The most prototypical acts of a category resemble

each other on a number of features or attributes that form the basis for the category and its definition (John, 1986). Finally, personality categories can be organized in the form of hierarchies (Hampson, John and Goldberg, 1986); they differ in their breadth or level of inclusiveness, with broader traits (e.g. *reliable*) subsuming more narrow ones (e.g. *punctual*).

This categorical conception is, in our opinion, ideally suited for the study of natural-language descriptors, because it provides a theoretical account of, and empirical procedures to examine, the fuzziness inherent in these terms. Despite the lack of clear defining boundaries, these terms can be extremely useful (Wittgenstein, 1953). Still, a more precise technical language for personality description is needed, and we have to begin somewhere. Even in animal taxonomy, as G. G. Simpson pointed out, 'the technical system evolved from the vernacular' (1961, pp. 12–13).

HISTORICAL BACKGROUND: EARLY TRAIT LISTINGS

In Figure 2 we give an overview of the history of personality-taxonomic research in the lexical tradition. In our review we will describe historical and unpublished material, which is generally difficult to obtain, in more detail than more recently published work. The English scientist and writer Francis Galton (1884) was probably the first who scanned a dictionary and assembled about 1,000 personality descriptors. Galton's work and that of other early investigators was relatively unsystematic and had little impact on the field. However, when Klages (1926) articulated the theoretical rationale for the lexical approach and argued that the study of language would benefit our understanding of personality, Franziska Baumgarten (1933) responded with a systematic study. To examine Klages' speculation that there were roughly 4,000 German words descriptive of 'inner states', Baumgarten assembled personality-descriptive terms both from various dictionaries and from the publications of German characterologists. Baumgarten selected only those terms that were most frequent according to her own judgement and did not classify them further. Descriptors of specific traits, such as *freundlich* (*friendly*) were listed along with purely evaluative words such as *schlecht* (*bad*) and *unwuerdig* (*unworthy*). In all, Baumgarten's list amounted to 941 trait-descriptive adjectives and 688 nouns, a number considerably lower than that estimated by Klages. As Figure 2 shows, only recently have some German taxonomers made use of this list (Angleitner, Ostendorf and John, in preparation). Well into the 1980s, the only list of trait-descriptive adjectives used for research in Germany had been a translation of Gough and Heilbrun (1965) *Adjective Check List*. However, Baumgarten's list did influence Allport and Odbert whose work has served as the empirical foundation for most later trait taxonomic research.

THE ALLPORT AND ODBERT LIST

To construct a listing of all personality-relevant terms in the English language, Allport and Odbert examined Webster's *New International Dictionary* (1925) which contained about 550,000 separate terms. Terms were included in the list if they were judged to possess 'the capacity . . . to distinguish the behavior of one human

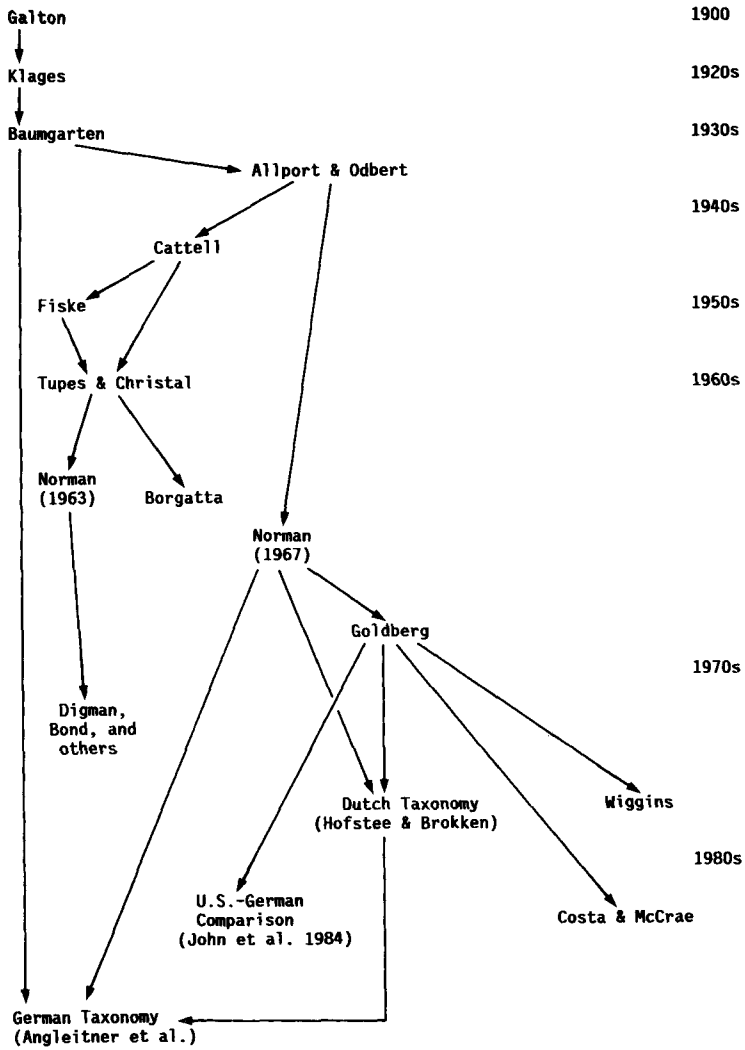


Figure 2. The history of lexical research on personality taxonomies

being from that of another' (Allport and Odbert, 1936, p. 24), a domain specification that eliminates terms that refer to common, non-distinctive behaviour. Allport and Odbert preferred adjectives and participles over nouns, which they included only if there were no corresponding adjectives or participles. With the addition of a few common slang terms not (yet) included in Webster's, the final list amounted to almost 18,000 words. Although Allport and Odbert listed only the most common form of a term if there were several variants, the number of terms included in their list was still of staggering size. In their attempt to be as inclusive as possible, even uncommon and derivative word forms were listed, and Allport and Odbert noted that 'probably half the terms seem strange and unfamiliar to the American reader' (p. 25).

The terms were listed in four categories or 'columns'. The first column contained 4,504 'neutral terms designating possible personal traits' (p. 38), which Allport and Odbert defined as 'generalized and personalized determining tendencies—consistent and stable modes of an individual's adjustment to his environment'. Obvious examples are *aggressive*, *introverted*, and *sociable* (p. 26). The definition of *stable traits* as internal and causal tendencies, as well as these examples, have guided most later taxonomic research.

'Terms primarily descriptive of temporary moods or activities' (p. 38) comprised the 4,541 terms in Column II. The majority of these words were present participles derived from verbs. Prototypical terms were *abashed*, *gibbering*, *rejoicing*, and *frantic*. The 5,226 terms listed in Column III were defined as 'weighted terms conveying social and characterial judgments of personal conduct, or designated influence on others', such as *insignificant* and *worthy* (p. 27). These terms referred primarily to social evaluations, rather than to neuropsychic dispositions that reside in the structure of personality itself. A number of these terms designated the 'social stimulus value' of a person, such as *dazzling* or *irritating*. Although these terms presuppose some traits in that person, they do not indicate the psychological dispositions that cause that person to have a dazzling or irritating effect on others. Column IV is a miscellaneous category consisting of at least four subcategories. The total of 3,682 words was labelled 'metaphorical and doubtful terms' (p. 38). One subset referred to physical qualities (e.g. *lean*, *redhead*), another to capacities and talents (e.g. *gifted*, *prolific*). Also listed in Column IV were terms that seemed to have doubtful relevance to personality, as well as those that could not be assigned to any of the three main categories.

Allport and Odbert noted that some of the words could have been classified into more than one column, especially those in Columns I and II. Therefore, some of their classifications had to be made rather arbitrarily. In an attempt to limit arbitrariness of classification, three independent judges each edited the entire list; the published version represents the final classification by Odbert. Unfortunately, each judge seemed to have 'a favorite column to which he assigns more terms than do the others' (p. 35). The mean agreement among the three judges about the classification of 300 representatively selected terms was 47%.

CATTELL'S MAJOR DIMENSIONS OF PERSONALITY

Cattell used Allport and Odbert's listing as a starting point for the development of an extensive multi-dimensional model of personality structure. Because Cattell's primary goal was to discover the major dimensions of personality lexicalized in English, his initial work was aimed at reducing the number of personality terms listed by Allport and Odbert to a more manageable size. This early work is of utmost importance because it has formed the foundation of Cattell's system of personality description, and provided the initial item selection for several other investigators. So far, little attention has been paid to this empirical work, probably because it is much less accessible than the global summaries in Cattell's various books.

Semantic reduction steps

Cattell's primary interest in stable traits led him to adopt Allport and Odbert's first

category, the personal trait terms, as the universe of personality-descriptive content. However, he added approximately 100 temporary-state terms, which he judged quite appropriate for personality descriptions (Cattell, 1943a, p. 488). The semantically similar terms on the expanded list were grouped as synonyms under a key word to reduce the number of personality characteristics that had to be examined separately. This semantic sorting task was carried out independently by Cattell (we presume) and a student of literature. Whereas Cattell (1943a, p. 488) reported that the two judges arrived at 'practically identical' lists of synonyms, he provided no quantitative information about interjudge agreement. Cattell's definition of synonymy is also not entirely clear. Whereas Cattell used the term *synonym* in a rather strict way in his writing, (i.e. terms that could be used interchangeably), the size of the synonym clusters suggests that a much looser criterion of synonymy was actually used; the number of synonyms ranged from 48 for *talkative* to 24 for *frank*.

Within each cluster Cattell added an opposite for each term. Terms that seemed to reflect dynamic (i.e. motivational) traits and ability traits were not paired with opposites because he viewed them as unipolar traits varying in intensity (e.g. from *not at all* intelligent to *very* intelligent). Moreover, for approximately 25 clusters, Cattell did not find any opposites and thus left them unipolar. Nevertheless, in the majority of clusters, the terms were listed as pairs consisting of bipolar traits. This format had the practical advantage of offering a more complete definition of the meaning of the cluster because it defined both extremes. However, the appropriateness of Cattell's intuitive pairings is difficult to evaluate.

The grouping into antonym pairs eliminated several clusters and permitted the classification of about 4,500 terms into 160 mostly bipolar clusters. To represent the clusters by a smaller number of terms, Cattell selected about 13 terms from each cluster and summarized them with a key term. Thus, at this stage Cattell had eliminated more than half of the terms listed by Allport and Odbert. Among those were all the words with prefixes whose word *stem* had already been included in a cluster; moreover, rare, obsolete, and strange terms were left out whenever their meaning was already represented by another term (Cattell, 1943a, p. 490).

To examine the exhaustiveness of his collection of personality variables, Cattell reviewed the personal literature available to him, particularly typologies and factor-analytic studies (e.g. Allport, 1937; Burt, 1937; McDougall, 1923; Sheldon and Stevens, 1942; Spoeherl, 1936; Spranger, 1914), as well as his own summaries of tests measuring temperament (Cattell, 1933, 1934) and intelligence (Cattell, 1936). Because Cattell found that only an emotionality factor established by Burt and two or three traits related to neurotic and psychotic disorders were missing from his 160 clusters, he concluded that his selection of variables was surprisingly complete. Nevertheless, to achieve a more elaborate representation of the behavioural domains captured by the clusters, Cattell supplemented some of his clusters with terms derived from the psychological literature; he also added the previously missing neurotic and psychotic terms.

However, Cattell found that interest and ability traits, already well-documented in the psychological literature in the 1930s and 1940s, were underrepresented in his set of clusters and specified in less detail than he deemed desirable. Cattell therefore added to his original 160 clusters (a) general intelligence (Spearman, 1927), (b) nine special abilities (i.e. manual dexterity, four factors derived from Thurstone,

drawing talent, musical talent, mechanical talent, physical strength, and endurance), and (c) 11 areas of interest (Spranger's six life forms, general aesthetic interest, interest in art and architecture, music, physical activity, family, and home). At some point Cattell must have eliminated a number of his initial clusters; otherwise it would be difficult to understand how the addition of 21 distinct *new* clusters to 160 existing ones could have led to a total of 171. In later publications (Cattell, 1957, p. 813; Cattell, 1979, p. 53; see also Sells, 1959, p. 4), Cattell referred to a trait list consisting of 181 or 182 clusters which, as he reports there, was the result of his semantic reduction steps. The published version of the 171 clusters (Cattell, 1946, p. 219–232) did not include all the synonyms for each cluster. Instead, only two to five terms were presented for each of the two poles in each cluster to exemplify the cluster name. This set was still much too large to be amenable to the factor-analytic techniques available in the 1940s. In his empirical analyses, Cattell thus proceeded in two steps. First, he used a clustering approach to condense the 171 clusters to a more manageable set of 35 'trait variables'. Then he submitted the intercorrelations among this reduced set of variables to factor analysis.

First empirical reduction: the clustering study

Cattell decided to use empirical data to guide his attempt to group the clusters into more inclusive variables, and thus obtained ratings of 100 adults. Every target person was rated on each of the 171 clusters by one 'intimate (but not emotionally involved) acquaintance' (Cattell, 1943a, p. 436), or by two associates (Cattell, 1945b, p. 140; 1946, p. 243); it is therefore difficult to ascertain the total number of raters employed in the study. The judges were instructed to *either* indicate whether the target person could be described better by the trait terms on the right-hand side of the cluster or by those on the left-hand side, *or* decide whether they viewed the person as falling above or below the average for that trait. Apparently, the former rating format was to be used for the bipolar clusters, and the latter for the unipolar clusters (most of which were interest, ability, and dynamic traits). The subjects must have found it difficult to use this list, given the inclusion of psychological jargon for abilities, interests, and components of neuroticism and psychoticism.

Using tetrachoric correlations, the 171 variables were correlated across the 100 target persons (an arduous task, one might add, given the computational facilities of the time). The resulting matrix of 14,535 correlations was laid out on paper covering an area of 14 square feet. Cattell then 'inspected' the matrix to identify a set of 30 to 40 representative variables that would contain as many of the initial 171 clusters as possible. To simplify this process, he limited the number of correlations he would consider, setting two arbitrary limits: (a) If two variables correlated at least ± 0.45 , they were considered to form a cluster, and (b) if they correlated at least ± 0.84 , they were considered identical. All of the 171 clusters that showed correlations exceeding either of these criteria were then listed. Cattell's further procedures have not been reported in sufficient detail to permit us to reconstruct the clustering process. We can refer only to Cattell's original description:

. . . every member correlates with every other member to the extent of 0.45 or more. . . . The variables with such correlation fell at first into two large groups of loosely tied (having many missing correlations) items and a third

less well-defined group. Closer inspection split these into small groups with none of the mutual correlations missing [i.e. none below 0.45]. The clusters having six to a dozen members were relatively few, but triads and tetrads [i.e. clusters of three or four] were quite numerous. This is not surprising when one reflects that a single cluster of 14 variables can be broken down into 364 separate triads. After excluding from the list any cluster which could be included in a larger cluster, we had left, in addition to the major clusters of six or more variables, approximately 15 pentads, 20 tetrads, and 88 triads (Cattell, 1943a, p. 499–504).

Subsequently, Cattell largely ignored the triads, considering only four of them significant enough to be retained. The final set included 67 clusters, although Cattell (1946, p. 245) later mentioned 69 clusters. Whatever the actual number of clusters at this stage, the published table (Cattell, 1943a, pp. 500–503) lists for each of 60 clusters a provisional cluster label, the triads forming the cluster core, and overlapping traits and clusters. Columns 4 and 5 of this table demonstrate a remarkable degree of overlap among the clusters in spite of the exclusion of correlations below 0.45. Yet, the table accounts for only 135 of the original 171 clusters. The missing 36 probably appeared to be of too little practical value to be retained as individual variables or were represented only among those 84 triads that were dropped from further analyses.

In conclusion, the clustering solution offered by Cattell seems neither well-formed nor exhaustive. Is it objective and replicable? Cattell claimed that 'the analysis into clusters was made entirely blind, on mathematical criteria only, and the experimenter became aware of the nature of the clusters only at the stage of listing, when it became necessary to give them provisional titles' (1943a, p. 504). However, Cattell did not describe the nature of his 'mathematical criteria' except that they were not as 'sensitive methods as Holzinger and Harman's B-coefficient technique or Tryon's system' (1943a, p. 499). The proposed cluster solution is therefore of dubious value. Given the number of undocumented subjective decisions, we do not see any compelling reason to expect that the same or even a similar solution would emerge even if the procedures *could* be replicated.

Incorporation of psychological findings and further revisions

Lack of funding for a factor analysis of all clusters compelled Cattell (1945a) to further reduce this set of clusters. For this purpose, Cattell (1945b; 1946, pp. 238–269) searched psychological publications for evidence of trait clusters. In addition to his own work, Cattell reviewed 12 other rating studies, and identified 131 trait clusters. He reported that he made all possible pairwise comparisons among these clusters, and combined those that overlapped. The resulting 50 clusters were subsequently arranged into 20 'personality sectors', most of which Cattell interpreted as well-known clinical syndromes. In fact, he considered these personality sectors important grouping principles for normal personality traits (e.g. Cattell, 1957). The first 12 of these sectors corresponded almost exactly to the 12 factors Cattell later identified in his factor analyses of trait ratings.

Cattell (1945a) used his summary of the psychological literature for a drastic revision of the cluster solution he had derived so painstakingly from his rating

study. In particular, of the 67 (or 69) 'empirical' clusters, he retained only 58 that had been confirmed by other studies. Moreover, when two or three 'empirical' clusters overlapped with one or several of the clusters derived from the work of other researchers, Cattell introduced a broader, more inclusive cluster. And, finally, another six 'less reliable' clusters were eliminated. When we added up all of the clusters—in part old, in part newly combined—39 should have remained that presumably had been confirmed by other investigators. According to Cattell (1945a), however, there were only 35. These 35 later became 36 and then 42 because at some point Cattell added 'a few terms considered indispensable' (1957, p. 813). In the original publication, the 35 variables each subsume six to 12 trait elements (Cattell, 1945a, pp. 71–74). All but two clusters have been arranged in bipolar form, suggesting that Cattell abandoned his earlier distinction between bipolar and unipolar traits.

In later descriptions of the 'same' variables (e.g. the 42 of 1957), Cattell augmented the trait adjectives by short descriptions, which changed the meanings of several clusters substantially. For example, Cluster 8 initially was labelled *rigid, tyrannical, vindictive vs adaptable, friendly* (1945a, p. 71). Later Cattell (1947; 1948) gave each variable a commonly used bipolar adjective pair as a heading, which was elaborated by phrases to define the meaning of the adjectives. In this version, the vindictive tyrant, central to the cluster in 1945, became a mere conservative. In 1957 (p. 813), the rigidity pole of that cluster read: '*Rigid*: Insists that things be done the way he has always done them; does not adapt his habits and ways of thinking to those of the group; nonplussed if his routine is upset', thus excluding both the hostility facet of the original cluster and the conservatism component added later on.

These changes were probably motivated by Cattell's realization that many of his initial cluster variables were multi-faceted and lacked coherence. Similarly, subsequent investigators have selected judiciously from the cluster descriptions. For example, whereas Cattell's (1945a) Cluster 35 contrasted 'sophisticated, intelligent, assertive' with 'simple, stupid, submissive', Fiske (1949), Tupes and Christal (1961), and most others used only the unidimensional contrast 'assertive vs submissive'.

The 12 primary personality factors

With the 35-variable list at hand, Cattell was ready to begin the final goal of his project—the identification of the major personality dimensions represented in language. For his first factorial study, Cattell (1945a) recruited 13 groups of 16 male adults each. The 16 subjects in each group were rank-ordered on each variable by two judges. If necessary, the judges were trained until their agreement correlation reached 0.60. Using the judges' mean rankings, the 35 variables were correlated in each group. The matrix representing the means of these 13 intercorrelation matrices then served as the input to the factor analysis. Cattell interpreted 12 obliquely rotated factors. However, he clearly retained too many factors; only the first three to five factors seem interpretable. The last three factors did not have even a single loading exceeding 0.30, and none of the factors beyond the fifth had loadings of variables that did not already appear on the first five.

In interpreting the findings from his subsequent studies, Cattell (1947, 1948)

concluded that he had replicated nine of the 12 original factors, which he then used as the foundation of his system of personality measurement and theory (e.g. Cattell, 1957). For example, to assess these factors in self-reports, he constructed the *Sixteen Personality Factors Questionnaire* (16PF); 12 of its 16 dimensions are said to match those found in the trait rating studies, whereas four are specific to the questionnaire domain. However, the high degree of correspondence Cattell claimed for his factors across self-report, ratings by acquaintances and intimates, and objective test data has been debated (e.g. Becker, 1960; Nowakowska, 1973). Moreover, reanalyses of Cattell's studies by others have not confirmed the number and nature of the factors he proposed (e.g. Digman and Takemoto-Chock, 1981); indeed, we know of no independent study of trait ratings that has replicated Cattell's 12 (or more) factors.

Five broad dimensions of personality description

Nevertheless, Cattell's work stimulated other researchers to examine the structure of trait ratings. Fiske (1949) constructed much simplified descriptions for 22 of Cattell's variables and used them to obtain trait ratings of 128 clinical-psychology trainees. The factor structures derived from self-ratings, ratings by other trainees, and ratings by the psychological staff were highly similar, suggesting five recurrent factors. Tupes and Christal (1961) reanalyzed correlation matrices from eight different samples, ranging from airmen with no more than high-school education to first-year graduate students who were rated either by peers, supervisors, teachers, or experienced clinicians in settings as diverse as military training courses and a sorority house. In all the analyses, Tupes and Christal found 'five relatively strong and recurrent factors and nothing more of any consequence' (1961, p. 14). They labelled their factors (I) Surgency (*talkative, assertive, energetic*), (II) Agreeableness (*good-natured, cooperative, trustful*), (III) Dependability (*conscientious, responsible, orderly*), (IV) Emotional Stability (*calm, not neurotic, not easily upset*), and (V) Culture (*intellectual/cultured, polished, independent-minded*). These factors resemble the first five in Cattell's (1945a) study, although that solution contained two separate Emotional Stability factors and combined Tupes and Christal's factors V and III into a single one. Five-factor structures similar to that described by Tupes and Christal have been reported by Norman (1963), Borgatta (1964), and Digman and Takemoto-Chock (1981). Because all their trait scales were derived from Cattell's 35 variables, these studies provide evidence that the relations among the variables in this set can be summarized by the five broad factors.

Cattell had argued that his factors represent the important dimensions of personality description, a claim based on his belief that his set of variables was indeed 'a truly representative list . . . derived from language' (1945a, p. 70). However, he conceded that 'naturally a factor analysis of these [35 variables] cannot be guaranteed to contain all the factors present among the original 171 traits of the personality sphere' (p. 71), and his concern with representativeness (or the lack thereof) is reflected in his later additions to the variable list. Unfortunately, as our historical detective work has shown, his various lists appear to represent primarily the traits Cattell himself considered the most important. Given that Cattell twice used his own literature reviews to evaluate and then revise and add psychological

jargon to the natural-language descriptors, the list cannot be representative of trait content in English. We admire Cattell's theoretical understanding and innovative genius; however, he did not concern himself with the details necessary to ensure the replicability of his empirical findings.

However, it is possible that Cattell's trait selection, although clearly not representative of the dictionary, may nevertheless contain clues to most of the major dimensions of personality description. Five-factor structures similar to those obtained with selections from Cattell's list have been identified subsequently in various sets of variables not directly derived from that list (e.g. Conley, 1985a; Digman, 1963, 1972; Digman and Inouye, 1986; Goldberg, 1980, 1981; McCrae and Costa, 1985, 1987). However, the scales used by Digman overlapped, in part, with those used by Tupes and Christal and by Norman (see Digman and Takemoto-Chock, 1981, p. 158). Similarly, McCrae and Costa, interested primarily in a reliable assessment of the five factors, used 40 scales selected by Goldberg as markers for the Big Five, and added another 40, which they wrote to assess each of the five dimensions as they interpreted them (i.e. including variables related to 'Openness to experience', their label for Factor V). Thus, it seems that there are only two studies whose results can be considered truly independent of the initial analyses of Cattell's variable list: In Conley's study the variables had been selected by Lowell Kelly even before the Allport-Odbert list was published, and Goldberg's research (reviewed below) was based on his own selection from the dictionary.

More recently, Peabody and Goldberg (1987) performed a comprehensive series of structural analyses and argued that one can think of the five factors as related to (I) Power, (II) Love, (III) Work, (IV) Affect, and (V) Intellect, respectively. Peabody and Goldberg used a set of variables selected with the goal of achieving an adequate representation of common English trait adjectives (Peabody, 1987). A comparison of their selection with that of Cattell showed that Cattell's 35 variables included markers for all the domains included by Peabody and Goldberg but did so in an unrepresentative manner. Compared to their selection, Cattell's variables seem to drastically *underrepresent* traits related to conscientiousness and impulse control (Factor III) as well as intelligence (a facet of Factor V), and *overrepresent* traits related to neuroticism (Factor IV) and culture (apparently only a minor facet of Factor V). The overrepresentation of emotional stability is probably due to Cattell's addition of clusters from the extensive literature on psychopathology, whereas the underrepresentation of intelligence can be traced to Cattell's use of an actual intelligence test to measure this characteristic in his later studies. If one assumes that Peabody and Goldberg's selection is indeed representative, it appears that Cattell's variable list was comprehensive enough to include at least a few markers for the very *broadest* of personality dimensions.

NORMAN'S 2,800 TRAIT DESCRIPTORS AND PRELIMINARY TAXONOMY

An exhaustive list of personality descriptive terms has been compiled by Norman (1967). The trait taxonomy he envisioned should be sufficiently exhaustive, precise, and well-structured to be useful for scientific communication, assessment, and the construction of personality theory. By *exhaustive* Norman meant that one should take 'as its fundamental data base the set of all perceptible variation in performance

and appearance between persons or within individuals over time and varying situations that are of sufficient social significance, of sufficiently widespread occurrence and of sufficient distinctiveness' (1967, p. 2). *Precision* would require that the final word list not contain terms with vague or ambiguous meanings. Finally, to develop a *well-structured* taxonomy, the 'relations between groups of terms are to be determined' (1967, pp. 2–3).

Exhaustive sampling of terms and subsequent reductions

To rectify the omissions and imperfections of the earlier work, Norman decided to begin anew with the second step outlined in Figure 1, identifying terms that would fit his specification of the domain of personality description. He used the unabridged 1961 *Webster's Third New International Dictionary* to identify terms that referred to attributes of persons or their behaviour but had not been included in the Allport–Odbert list. However, most of the 9,046 additional terms he found in *Webster's Third* were simply suffixal and prefixal variations of terms already included. Thus, Norman added only 171 terms, resulting in a master set of 18,125.

To eliminate unsuitable terms from this initial set, Norman used the four categories that are listed, along with some samples, in Table 1. The first exclusion

Table 1. Norman's (1967) domains of personality description and his four exclusion categories

Category number and label	Number of terms in category	Examples
Stable 'Biophysical' Traits		
Cat. 1: Prime trait terms	608 (3%)	calm, helpful
Cat. 2: Difficult trait terms	544 (3%)	abtuse, pedantic
Cat. 3: Slangy or quaint trait terms	1645 (9%)	chirpy, icy
Subtotal:	2797 (15%)	
Temporary States and Activities		
Cat. 4: Prime terms: States	384 (2%)	hesitant, sad
Cat. 5: Prime terms: Activities	583 (3%)	bickering, quitting
Cat. 6: Difficult terms	399 (2%)	euphoric, livid
Cat. 7: Slangy or quaint terms	1655 (9%)	bawling, smitten
Subtotal:	3021 (17%)	
Social Roles, Relationships, and Effects		
Cat. 8: Prime terms: Roles	242 (1%)	leader, subservient
Cat. 9: Prime terms: Effects	163 (1%)	attractive, humorous
Cat. 10: Difficult terms	163 (1%)	droll, enigmatic
Cat. 11: Slangy or quaint terms	908 (5%)	appetizing, repellent
Subtotal:	1476 (8%)	
Included Categories Subtotal:	7294 (40%)	
Excluded Categories		
Cat. 12: Evaluative terms	760 (4%)	capable, nice
Cat. 13: Anatomical, physical terms	882 (5%)	good-looking, healthy
Cat. 14: Ambiguous and vague terms	4796 (27%)	aesthetic, mannered
Cat. 15: Obscure terms	3606 (20%)	eldritch, scait
Excluded trait terms (too difficult)	787 (4%)	rattle-brained
Subtotal:	10,831 (60%)	
TOTAL:	18,125 (100%)	

category (Cat. 12) contained evaluative terms and mere quantifiers. These terms tend to be seen either as almost exclusively evaluative (i.e. descriptive of social and personal approval) or as merely modifying the degree or amount of an attribute (e.g. *awful, nice, gifted*). Another exclusion category (Cat. 14) contained ambiguous, vague, and tenuously metaphorical terms. The meaning of these terms can be interpreted in several different ways. Also included here were terms that refer to very broad classes of behaviour which have few, if any, attributes in common, as well as those terms whose meaning could be understood only by the use of analogies or metaphors. Category 15 comprised very difficult, obscure, and little-known terms. Many of these words are not used in contemporary English, or have obscure literary, historical, or mythological references.

Whereas these three categories are all based on the characteristics of the words themselves, Norman also excluded terms that refer to one particular domain of personality. Category 13 combines anatomical and physical dispositions and conditions. Included here were all the terms that denote physical and mental health (e.g. *ill, senile, insane*), physical dispositions (*athletic, short*) and current conditions (*jogging, obese*). Also listed in this category are those aspects of physical attractiveness (*handsome, pretty*) that do not primarily denote social effects (*arousing, sexy*). In defense of Norman's exclusion of these domains, one might argue that physical health and attractiveness are not personality characteristics *per se* but important influences *on* personality. However, because the reverse causal link (from personality to health and attractiveness) may be also important, these characteristics should be included in an exhaustive specification of personality. In all, the application of these exclusion criteria reduced the initial set by more than 50 percent, although the criteria were applied quite conservatively; three of the four team members had to agree before a term was assigned to one of the four categories.

Domains of personality description

The remaining 8,081 terms were then sorted into three major domains: (1) stable 'biophysical' traits, (2) temporary states and activities, and (3) social roles, relationships, and effects. Each of these categories was subdivided into (a) 'prime' terms, (b) moderately difficult terms, and (c) 'slangy', quaint, awkward, and colloquial terms. The resulting category system, summarized in Table 1, represents an elaborate, though implicit, conception of personality.

Whereas Norman concentrated his subsequent *empirical* research on the stable-trait category, his view of personality was considerably broader. The best way to appreciate the conceptual distinctions embodied in Table 1 is to consider one particular characteristic across the categories. In particular, an individual's personality may be described by an enduring trait (e.g. *irascible*), by the internal states the individual typically experiences (*furious*), by the activities the individual frequently engages in (*yelling*), by the effects the person has on others (*frightening*), or by a general evaluation (*terrible*). Norman classified the term *irascible* as referring to a stable trait (Cat. 1-3). This disposition, often called anger proneness or 'trait anger', may be manifested, at any given time, in the individual's *temporary* conditions. Norman differentiated two kinds of such conditions, states (Cat. 4) and activities (Cat. 5). *Furious*, for example, is a prototypical state

descriptor, whereas *yelling* is a prototypical activity descriptor (Chaplin, John and Goldberg, 1988). Although the same trait may give rise to both kinds of temporary conditions, the concepts State and Activity are distinct from the concept Trait.

The fourth class of descriptors defined by Norman (Cat. 8 and 9) also seems causally linked to those in the trait category. Terms such as *frightening* and *intimidating* do not refer to traits but to the effects individuals (particularly *irascible* ones) have on others; these terms signify an individual's 'social stimulus value' or reputation (Allport and Odbert, 1936). The mediating link between trait and social stimulus values are the individual's behaviours and internal states because they reflect the trait and constitute the stimulus that influences others in the individual's social environment. Finally, the natural language provides additional descriptors that are even more evaluative than are terms like *frightening*, for example, *horrid*, *terrible*, or simply *bad*. Norman argued, following Allport and Odbert, that these terms contained too much evaluation and too little descriptive meaning to be useful for descriptive purposes, and thus assigned them to an exclusion category (Cat. 12).

In conclusion, our analysis of Norman's category scheme suggests that he endorsed a broad conception of personality but saw traits as the central concepts. It appears that he viewed traits as enduring characteristics internal to the individual and causally effective, a view quite similar to that of Allport (1937). However, Norman excluded from this dispositional conception of traits those stable dispositions that are related to physique and health, a distinction that implies the concepts of *disposition* and *trait* should not be equated because the former is considerably broader.

Towards a well-structured taxonomy

Norman's (1967) listing provided the foundation for most contemporary taxonomies because the exclusion and inclusion of terms was based on explicit criteria and the consensus of a team of four judges. Moreover, Norman subsequently accumulated an impressive array of rating data on a set of almost 2,800 trait terms. Fourteen lists of 200 words were each given to a different sample of 50 male and 50 female undergraduate students. They were instructed to write down a synonym or short definition for the term or to cross it out if they did not know its meaning. Then they judged the degree to which each word was descriptive of themselves, and of a well-liked peer, a neutral peer, and a disliked peer. Finally, they rated the social desirability of each term. Analyses of these data showed that the familiarity, currency, and specificity of meaning of many terms had to be considered doubtful. About 1,200 terms were thus excluded, on the basis of their high level of difficulty, their extremity in the self-ratings, or their slanginess. The remaining 1,566 trait terms thus constituted the instances to be included in his taxonomy.

This extensive pool of trait descriptors has been used by Norman to examine two central issues. First, although in 1963 Norman had published his by now classic replication of the five factor structure (which we will refer to as the 'Big Five'), all the analyses published so far had relied on Cattell's variable set. Would there be only those five 'and nothing more of any consequence' if a more extensive and representative pool were used? Or would additional, or maybe even entirely different, factors emerge? The second issue centres on the exact nature and composition of these broad factors. The Big Five are too broad for many purposes

of personality description. Biologists, in contrast, have long recognized that it is 'absolutely necessary in scientific systematics to have a system in which different levels of generality or inclusion are recognized' (Simpson, 1961, p. 12). Similarly, a complete trait taxonomy must include middle-level categories, such as Dominance, Orderliness, and Creativity, as well as even more narrow and subordinate ones, such as *talkative*, *punctual*, and *musical* (Hampson *et al.*, 1986).

The first of these issues (i.e. are five factors sufficient to represent the domain at the highest level of abstraction or breadth?) led Norman and later Goldberg to search for groups of terms that could not be assigned to one or another of the Big Five dimensions. The second issue (i.e. what are the components subsumed by the broadest factors and how can these components be subdivided even further?) led Norman to develop a preliminary hierarchical structure for the entire domain of trait terms.

Norman's preliminary hierarchical classification

In a first step, Norman (see Goldberg, 1980, 1981) rationally sorted his almost 1,600 terms into 10 broad classes, one for each of the two poles of each of the Big Five dimensions. Using the findings from the previous factor analyses as a guide, Norman had little difficulty assigning almost all of the terms to a factor pole. However, whereas Norman (1963) had selected an equal number of markers for each of the Big Five in his factor analyses, his sorting showed that the Big Five do not represent equally 'big' domains of English trait descriptors. The number of terms per factor and pole varied considerably, ranging from a low of 64 for IV (Neuroticism pole) to 274 for II (Disagreeableness pole), with an average of 155 terms. These findings are consistent with Peabody and Goldberg's (1987) later (and independent) observation that, when a representative selection is factored, IV and V tend to be the smallest and II by far the largest factor.

Norman then sorted the terms within each of the 10 factor poles into more narrow semantic categories. The sorting of the terms at the Neuroticism pole of Factor IV led to only three distinct categories, labelled Anxiety, Insecurity, and Self-pity, whereas 12 categories emerged at the Disagreeableness pole of Factor II (Vindictiveness, Ill-humor, Criticism, Disdain, Aggressiveness, Antagonism, Dogmatism, Temper, Distrust, Greed, Callousness, and Uncooperation). In total there were 75 such middle-level categories, most of which still contained a sizeable number of terms. For example, Dogmatism consisted of *biased*, *opinionated*, *stubborn*, *inflexible*, and 45 other related terms. Norman therefore examined the semantic relations among the terms within each of the 75 categories, and combined highly synonymous terms into sets, a procedure which led to a total of 571 synonym sets. This three-tiered hierarchical structuring allowed Norman to classify a total of 1,431 adjectives and 175 nouns, leaving only 25 words unclassified.

The 75 categories at the middle level of abstraction in Norman's taxonomy provide a compromise between the parsimonious but relatively undifferentiated superordinate level (10 factor poles) and the highly differentiated but unwieldy number of synonym sets at the most specific level. Moreover, the hierarchical nature of this classification provides the user with a quick and efficient elaboration of each middle-level category (by examining the synonym clusters included) as well as with a more general appraisal of its gist (by referring to the Big Five pole under

which it is subsumed). The heuristic utility of this classification is illustrated best with Factor V, whose various interpretations (e.g. as Culture, Intellect, and Openness to experience) have been the subject of much recent debate (e.g. McCrae and Costa, 1985). Norman's conception of this factor becomes more transparent than in his 1963 article when one considers the following middle-level categories (each elaborated by a synonym cluster): Formality (*pompous*), Grace (*dignified*), Vanity (*affected*), Sophistication (*urbane*), Maturity (*mature*), Wisdom (*intelligent, philosophical*), Originality (*creative*), Knowledge (*informed*), and Art (*artistic*), vs. Provinciality (*unrefined, earthy*), Imperceptiveness (*ignorant, narrow*), and Immaturity (*naive, superstitious*). Of course, this view of Factor V represents that of a single investigator, and others will disagree with some or most of the specifics.

In conclusion, Norman (1963) began by selecting a subset of 20 variables from Cattell's set that clearly defined the Big Five in factor analyses. Norman's subsequent hierarchical taxonomy included the comprehensive set of trait descriptors he had constructed from the dictionary; in his procedures, he was guided initially by his interpretation of these factors and later by the semantic similarity among the terms in each of the domains demarcated by the factors. Thus, at the top level, Norman's hierarchical classification is constrained by a selection from Cattell's limited variable selection, whereas at the lowest level it contains a comprehensive sample of trait descriptors grouped by their semantic similarity.

THE GOLDBERG TAXONOMIES

The robustness of the Big Five—Plus a Little Two?

Goldberg (1980, 1981, 1982) continued where Norman had left off. One of his goals was an empirical examination of Norman's preliminary classification. For this purpose, Goldberg constructed an inventory of 1,710 trait terms that could be used in studies of self and peer descriptions. On the basis of Norman's work, Goldberg (1982) excluded from the complete set of 2,797 stable-trait terms the 232 nouns, as well as 250 adjectives whose meanings were unknown to more than a third of Norman's student sample, and 650 adjectives which the research team judged as slangy or awkward variants of more common terms already included. Goldberg then compared this list with independently constructed ones by Anderson (1968) and Gough and Heilbrun (1965). As a result, he added 44 terms to permit comparisons among instruments in later empirical studies.

Among the 1,710 trait adjectives were 1,431 which Norman had classified into 75 categories. Based on the self-ratings of 187 college students using these terms, Goldberg (1981, pp. 159–161) analyzed the correlations among 75 category-scale scores formed on the basis of the terms included in each category. When five factors were extracted, the by now familiar Big Five emerged across a variety of different methods of factor extraction and rotation. The robustness of this factor structure is due, in part, to the fact that Goldberg did not rely on single adjectives but on scores which were each based on a large number of single adjectives and had Alpha reliabilities in the range of 0.70 and above. Another interpretation is that, because the construction of the 75 categories had been guided by the Big Five, the emergence of these factors is not very surprising.

However, when more than five factors were rotated, additional factors were found. When six factors were rotated, the categories Norman had classified into Factor V split apart into (a) an *Ability* factor (Wisdom, Originality, Knowledge, Art, plus Objectivity and Reflection which Norman had assigned to Emotional Stability; Imperceptiveness loaded negatively) and (b) a *Culture* factor which retained the other Factor V categories (except Maturity which loaded on Conscientiousness) plus several categories related to religiosity. In the seven-factor solution, the latter categories (Religiosity, Evangelism, Passionlessness, plus Honesty, vs. Irreverence) formed a small factor of their own.

These as yet unpublished findings suggest that the 1,710 trait adjectives, as summarized in the 75 categories, include broad dimensions *in addition* to those represented in Cattell's 35-variable selection. The two additional dimensions have some parallels in other studies. For example, on the basis of reanalyses of six studies, Digman and Takemoto-Chock (1981) interpreted their recurrent Factor V as Intellect, and a less stable sixth factor as Culture (or Social Class). The religiosity scales have, to our knowledge, not been used in an inventory besides the 1,710. However, Peabody and Goldberg (1987) found some evidence for a small sixth factor which they interpreted as *Values* (*honest, moral, fair*). Given that these terms were included in Norman's Honesty category which loaded the Religiosity factor in the 1,710, both factors may reflect the same underlying characteristic.

The 1,710 still included many terms with which even college students are unfamiliar. In subsequent studies, Goldberg (1982) used various abbreviated sets of more commonly-used trait terms, which he culled on the basis of various criteria, such as difficulty, ambiguity, slanginess, and sex-linkage. In a study of one such list, Goldberg obtained ratings of 475 very common trait adjectives which had been sorted (without guidance from the Big Five) into 131 sets of 'tight synonym' clusters. In two samples of peer ratings and of self-ratings, the seven-factor structures were very similar to each other and to the structure obtained from Norman's categorization of the more exhaustive 1,710. It thus seems safe to conclude that, if Factor V is interpreted as Intellect or Ability, the two prime candidates for an extension of the Big Five are Culture and a Value-Morality factor. Nonetheless, even a 'Big Five Plus Little Two' seems too limited to provide an adequately differentiated description of an individual. A descriptive level of lesser breadth was still needed.

Middle-level categories: the evaluation-explicit taxonomy

In the course of these factor analytic studies, Goldberg noted some deficiencies in Norman's 75 middle-level categories and in the synonym sets they subsumed. Norman had begun his classification with factor poles that are best thought of as overlapping sets with fuzzy boundaries. However, in his classification Norman treated the factor poles as if they were mutually exclusive; as a consequence, some semantically similar terms were assigned to different factors. Moreover, since each term was considered singly, the two terms in root-negation pairs (e.g. *emotional-unemotional*) were not necessarily assigned to the same factor. In contrast, when constructing his taxonomy Goldberg (1982) made explicit use of the antonymic relationships among the terms and represented the same number of terms as Norman in about 40 bipolar categories.

To make the construction of his semantic categories as systematic as possible, Goldberg employed a particular sorting algorithm. This algorithm was based on a scheme that Peabody (1967, 1968, 1970) had developed to unconfound denotative and connotative components in the meanings of trait adjectives. In this scheme, a bipolar descriptive contrast (e.g. saving vs spending money) is considered at various levels of evaluation (or desirability), such as *thrifty* vs *generous* (both desirable) and *stingy* vs *extravagant* (both undesirable). Goldberg used this scheme to sort the terms into categories that specified not only the descriptive contrast captured by the category but also the evaluation (i.e. mean social desirability ratings) associated with each term.

Because he was concerned with comprehensiveness in the initial stages (or 'rounds') of his categorization, Goldberg began with the 1,566 terms on Norman's trait list, which he sorted into 47 'evaluation-explicit' semantic categories; approximately 100 terms remained uncategorized. In his next rounds, Goldberg slightly modified and expanded Norman's set of terms to the 1,710 described above. The 45 semantic categories at that stage (Round V, see Goldberg, 1982, p. 225) were treated as scales and scored for the 187 subjects who had participated in the afore-mentioned study of self-ratings using the 1,710. Analyses of the internal consistency of each category-based scale, and factor analyses of their intercorrelations, led to several changes in the initially purely semantic classification. The appropriateness of these changes was then again evaluated in the self-rating data.

In later rounds of his taxonomy, Goldberg focused on smaller sets of terms that are more easily understood by college-educated native speakers of English, to permit the collection of extensive empirical data on these terms. Moreover, to facilitate the application of the evaluation-explicit sorting scheme, Goldberg included 34 amplifications (e.g. *overfriendly*, *overgenerous*), none of which was listed in the unabridged dictionary, plus two negations not included in Norman's list and one root term (*contemptuous*) that had been removed in the reduction of Norman's list to Goldberg's 1,710. Finally, he added 24 terms, primarily from Norman's categories of temporary states and ambiguous terms. After several further iterative cycles of semantic sorting and empirical evaluation and revision, the last round of Goldberg's taxonomy included 893 adjectives in 42 categories.

Among the 42 categories included in the final version, there were four (i.e. Conservatism vs progressiveness, Knowledge and education, Religiosity, and Femininity vs masculinity) that according to Goldberg (1982, p. 226) 'are not strictly personality traits'. Although they had been included in the 1,710 (probably added to permit comparisons with other inventories), Goldberg did not view religious and political attitudes, social roles, effects, and sexuality as *proper* traits. This view of traits might also explain why Goldberg did not emphasize the small Values-Religiosity factor he had identified in his earlier factor analyses.

Taxonomies of other personality domains

Whereas most previous taxonomers have restricted themselves exclusively to trait terms, Goldberg (1982) and his research team also constructed preliminary taxonomies for two other domains. One taxonomy covered the 384 terms in Norman's fourth category, prime terms for temporary states, moods, and attitudes. These terms were grouped according to semantic similarity into 13 categories, three

of which were further divided into subcategories. Included among these categories were most, if not all, of the basic emotions postulated by emotion theorists, plus Gratitude, Commitment, and a small category of Physical conditions, such as *drunken* and *hungry*.

Another preliminary taxonomy covered the prime terms of Norman's Categories 8 and 9, Social roles and effects, supplemented by a few of Norman's categories of Difficult terms and of Slangy or quaint terms for these domains. This is the taxonomy where one finds those terms that refer to an individual's social reputation and success, such as Fame, Competence, Respect, and Popularity. Also included here as separate categories are aspects of an individual's social stimulus value, such as Attractiveness, Charm, and Entertainment value, as well as attitudes towards Politics, Religion, and Sex, which had been represented in the trait taxonomy by only a few terms. Thus, whereas it may appear that Goldberg ignored these individual differences, in fact, he simply classified them elsewhere.

Comparisons with another source of personality terms: nouns

Whereas some personality nouns, such as *niggard* and *sucker*, had been included in the Allport–Odbert and Norman lists along with the adjectives, Goldberg and his co-workers treated nouns as a separate word class. To develop a preliminary taxonomy of these nouns, two abridged dictionaries were scanned for relevant terms, and the resulting list was supplemented according to informal suggestions. The 1,342 most common of a total of 1,947 nouns were classified into 50 categories; 42 were those previously developed for the trait adjectives, and only eight (one of which was a Miscellaneous category) had to be added to account for the 493 nouns that could not be assigned to the adjectival categories. Because the nouns had been neither classified into descriptive domains, nor culled and refined as extensively as the adjective sets, the seven noun-specific categories offer some glimpses of characteristics that might be relevant to personality but had been excluded from previous adjective sets.

The noun categories Attractiveness, Strength and size, and Awkwardness vs coordination denote physical appearance and motor behaviour and thus belong to one of Norman's exclusion categories (Cat. 13). The categories Youth vs maturity and Popularity and success reflect Social roles and effects (Norman's Cat. 8 and 9), and correspond to categories in Goldberg's taxonomy of that domain. Finally, Hedonism and Conformity seem to refer to stable and enduring characteristics that should be included among the traits. A close inspection of Goldberg's adjective taxonomy showed that terms related to Conformity were included in the Assertion vs cooperation category. However, it might be useful to distinguish genuine cooperation (i.e. behaviour motivated by concern for others) from conformity (i.e. behaviour motivated by fear of social sanctions). Indeed, Conformity, as contrasted with Autonomy and Independence, has been difficult to place in the traditional Big Five structure. Although usually classified along with Factor I scales, such as *submissive* and *timid* (John *et al.*, 1984), ratings of conformity and independence tend to have stronger (though still moderate) loadings on Factor V, a finding which is consistent with McCrae and Costa's (1985) interpretation of that factor as Openness to experience. The other noun-specific trait category, Hedonism (as distinct from Selfishness), does not have a direct equivalent among

Goldberg's adjective categories. However, to the extent that Hedonism contrasts with traits that involve self-restraint motivated by *moral*, *ethical*, or *religious* considerations (see Goldberg's adjective categories for Religiosity and Conservatism, as contrasted with Passion), the Hedonism category may reflect the Values factor we described earlier.

In conclusion, with the exception of Conformity and Hedonism the nouns yielded categories quite similar to those represented in Goldberg's three adjective-based taxonomies, with the structures resulting from the classifications of adjectives being more elaborate and differentiated. Indeed, Goldberg noted that the number of personality-descriptive adjectives far exceeded the number of nouns. Moreover, a higher percentage of nouns were associated with extremely negative evaluations and were judged as slangy or colloquial than were the adjectives. If one assumes that extreme evaluations and slang primarily serve an emphatic-expressive function in oral discourse, then one would expect that nouns, as compared to adjectives, are used more frequently in oral discourse than in writing.

WIGGINS' CIRCUMPLEX TAXONOMY OF INTERPERSONAL TRAITS

Because the circumplex taxonomy has been described and reviewed in several succinct and recent publications (e.g. Wiggins, 1979, 1980; also Buss and Craik, 1983), we will here focus on the assumptions that are unique to this taxonomy and thus elucidate corresponding assumptions that are often left implicit in other taxonomies. Considering again the process of taxonomy construction outlined in Figure 1, Wiggins' taxonomy differs from Norman's and Goldberg's primarily in its inclusiveness and in the strategies used to structure its domain. With respect to inclusiveness, Wiggins began with Goldberg's list of 1,710 trait adjectives but then divided that domain further into six subdomains—interpersonal traits (e.g. *aggressive*), material traits (*miserly*), temperamental traits (*lively*), social roles (*ceremonious*), character (*dishonest*), and mental predicates (*analytical*)—and limited his taxonomy to one such subdomain, the interpersonal traits.

With respect to the structural principles, Norman had chosen a hierarchical model, whereas Wiggins adopted Foa and Foa's (1974) circumplex model of interpersonal behaviour as a theoretical framework. In general, a circumplex requires a circular arrangement of variables in a two-dimensional space. For their interpersonal circumplex, Foa and Foa chose two broad dimensions, Status (or Dominance) and Love (or Warmth), which seem to be universally applicable to human interaction. Wiggins adopted these two dimensions as the major axes of his circumplex, so that the relations between one interpersonal trait and any other can be specified with reference to these two axes. A final, procedural difference, namely the use of a deductive rather than an inductive structuring algorithm, was also a result of Wiggins' theoretical perspective. Cattell and Goldberg had both aimed at *discovering* a structure and therefore used sorting and factoring procedures from which the taxonomy emerged gradually. Wiggins, in contrast, had already selected an *a priori* structure and needed to identify terms whose properties fit the constraints imposed by that structure.

Wiggins (1979) included in his specification of the interpersonal domain those 817 of the 1,710 trait adjectives that, in his team's judgement, referred to

interpersonal behaviours ('what people do to each other'). Based on correlational analyses of self-ratings on these traits, the terms were assigned to 16 categories, arranged in the form of a circumplex. The number of terms initially assigned to each category varied considerably, ranging from 29 to 95. Successive refinements of these categories eventually led to 16 scales, each consisting of eight single adjectives. These scales quite closely approximate the postulated circular structure and provide a comprehensive framework which has brought some order and integration to the vast array of models and measures in the domain of interpersonal behaviour (Wiggins and Broughton, 1985).

THE DUTCH TRAIT TAXONOMY

As shown in Figure 2, most comprehensive taxonomies have been constructed in English, primarily because the taxonomers were American. Although a German taxonomy project is now well under way (Angleitner *et al.*, in preparation), the only taxonomy not based on English is the one developed by Hofstee and Brokken at the University of Groningen in the Netherlands (Brokken, 1978). One of the foremost concerns of the Dutch team was that taxonomy construction is often a contentious process. Personality taxonomers decide, for example, whether a particular term is relevant to personality description, whether it refers to a stable trait or not, and whether it is too vague or ambiguous to be useful. They also interpret the psychological meaning of a particular category or factor position. All of these decisions require judgement; ideally, such judgement is based on an explicit conceptual rationale which is applied in a consensual way, rather than on the vague and often subjective intuitions of a single investigator. The Dutch team tried to avoid as much as possible the need for subjective decisions of the kind that had marred Cattell's taxonomic efforts. Among the most important of their many methodological contributions were procedures to (a) objectify the process by which instances that fit the domain specification are being identified, (b) ensure the generalizability of the resulting structure across judges and data sources, and (c) aid the interpretation of structural categories and factors by independently obtained consensual data.

Identifying instances: what is a personality-descriptive adjective?

The Dutch team used the 1970 edition of *Van Dale's Unabridged Dutch Dictionary* as their source. Two research assistants independently culled all adjectives except those that in their opinion could not possibly apply to a person. Their combined list included 8,690 adjectives. Four other team members then made a preliminary screening of the list, excluding mistakes, abstruse terms and jargon, metaphorical and purely evaluative terms, anatomical, physical or medical terms, as well as terms used only for temporary states or moods (Hofstee, 1976, p. 5). Obviously, these initial exclusion criteria were inspired by Norman's (1967) work. The four judges each had rejection rates of around 50 per cent and their classifications (reject vs not) generally matched. Nevertheless, to ensure comprehensiveness, only the 2,635 terms about which the four judges agreed unanimously were discarded. Paid judges were then recruited to classify the remaining 6,055 adjectives into two categories,

those descriptive of personality and those not descriptive of personality. In order to define the concept 'personality descriptive' for the raters, the Dutch taxonomy team developed two operational definitions, each phrased as an intuitively meaningful and heuristic criterion sentence that 'would act like a sieve' and retain only the terms relevant to personality (Brokken, 1978, p. 17).

The first of these, the *Nature* criterion, read: 'Call an adjective a "personality descriptive adjective" if it may replace the dots in the following sentence: He (She) is . . . by nature' (Brokken, p. 17). The Dutch wording for 'by nature' was 'van aard' (in German, 'von Art'). The terms were classified by two samples of graduate students, 10 of whom studied psychology and eight studied Dutch. The classification percentages derived from the two samples correlated 0.84 across the 6,055 adjectives, and interjudge agreement was quite high in both samples (mean Alpha = 0.85). Not surprisingly, the majority of the adjectives were not considered personality-descriptive and thus received low scores; 1,771 were rejected by all 18 judges whereas only 41 were accepted unanimously.

An inspection of the terms judged personality-descriptive according to the *Nature* criterion indicated that many adjectives commonly used in everyday descriptions of oneself and others would have to be excluded if only that criterion were used; apparently, the *Nature* criterion favoured the selection of temperamental traits over other types of traits. 'In Dutch, however, "van aard" is not necessarily genetic or physiological, even though it has temperamental associations (Hofstee, personal communication, Sept. 18, 1986). To examine the convergent validity of this operational definition, and to avoid a narrowing of the personality domain to temperament traits, a second criterion was formulated: 'Call an adjective "personality descriptive" if it can be used in answering a question like: What kind of person is Mr. (Mrs.) X?' Using this *Person* criterion, 25 psychology students classified each of the 6,055 terms. Again, the classification percentages were highly reliable. Moreover, the correlation between an adjective's *Nature* and *Person* classification percentages approached 0.90, an impressive demonstration of the convergent validity of these two operationalizations.

Overall, the classification percentages on the *Person* criterion were somewhat higher than those on the *Nature* criterion; apparently the *Person* criterion was a 'wider sieve' defining personality somewhat more broadly than did the *Nature* criterion. To understand the conceptual meaning of the two criteria better, we examined the terms with high *Person* but low *Nature* classification percentages which are listed in Brokken (1978). Apparently, the *Person* criterion elicited higher classification percentages for terms that refer to (1) a person's effect on others (e.g. *adorable*, *fearsome*), (2) social evaluations (e.g. *banal*, *capable*, *senile*), and (3) appearance (e.g. *delicate*, *pretty*), domains that none of the previous taxonomers had classified as stable traits. In contrast, the relatively few terms receiving high *Nature* but low *Person* classifications, such as *egocentric*, *overambitious*, and *rigid*, seem to correspond quite well to earlier notions of traits. Moreover, confirming Hofstee's interpretation of 'van aard', the characteristics to which these terms refer are not obviously biologically-based but refer to character traits based, at least in part, on learning.

In a third study, subjects rated the degree to which they saw each term as a descriptor of a 'fundamental' personality characteristic. If people consider traits as more fundamental to personality than social effects and appearance characteristics,

then the Nature classifications should be correlated more strongly with the Fundamentality ratings than should the Person classifications; in fact, this pattern of correlations was found. This finding also provides some consensual validation for Norman's (1967) decision to eliminate both appearance-related and purely evaluative terms and to set aside social effects until the more 'fundamental' class of traits had been taxonomized adequately. Given the high correlation between Nature and Person scores, the two criteria were combined to maximize the reliability of the final culling decisions; all terms that reached a classification percentage of at least 50 percent on the combined criterion were retained. The resulting set of 1,203 adjectives is listed in Brokken (1978, pp. 96–109).

Major dimensions of trait description in Dutch

Subsequently, the Dutch team focused on the identification of the major dimensions of trait description represented in their set of characteristics. In a study of 200 pairs of subjects who described either themselves or their friend, Brokken (1978) factored the correlations among all 1,203 terms based on either the self or the peer ratings. He then rotated six factors extracted from each of the two data sets to a position of maximum congruence. To aid interpretation, the loadings of the terms on each factor were correlated with independent ratings of the terms on Osgood's dimensions of Evaluation, Activity, and Potency. The loadings on three factors (labelled Dominance, Sprightliness, and Extraversion) were associated with the Activity ratings, suggesting that they are all related to the more inclusive Surgency factor in the Big Five. In addition to these associations with Activity, the factor labelled Dominance was associated even more strongly with the Potency ratings, whereas Sprightliness was somewhat associated with positive Evaluation, suggesting that this factor reflects not only activity level but also positive affect (e.g. *cheerful* and *sunny* vs *apathetic* and *depressed*). The first factor, strongly associated with positive Evaluation, could be identified clearly as Agreeableness (or Pleasantness). The remaining two factors were not associated with the Osgood dimensions. One was clearly Conscientiousness (*orderly* and *precise* vs *disorganized*), whereas the other combined terms for Emotional Instability with those Factor V scales that McCrae and Costa have interpreted as Openness to new experience and change (e.g. *complex*, *creative*, *nonconforming*, *progressive*, *rebellious*, and *undogmatic*). Although this factor would make sense given that the subjects were Dutch students in the early seventies, the enormous number of 1,203 variables probably led to low reliabilities and to the relatively small loadings found in this analysis. It thus seemed desirable to repeat the analysis after combining the ratings of several terms into more reliable scales.

For this purpose, Hofstee (1977) used the categories in Goldberg's trait taxonomy to sort the 1,203 Dutch adjectives into 42 clusters. Separate factor analyses of the adjectives in each of these clusters yielded 143 'scales', each represented by two to five adjectives. This set was reduced by removing unreliable and overlapping scales, and then used to construct 96 clusters consisting of two or three synonyms each. The definition of synonymy was not left to the semantic intuitions of the investigator but determined empirically, using the factor loadings and the previously obtained ratings of Activity, Potency, and Evaluation. A factor analysis of the intercorrelations among the 96 clusters yielded seven factors, five of

which corresponded to the Big Five. The two additional factors were labelled Aggression (Irritation) and Conservatism. To facilitate further investigations in Dutch, and to relate their seven adjective factors to other measures of personality, the Dutch team published a 'Standard Personality-Trait List' (Hofstee, Brokken and Land, 1981) which has since become a standard personality instrument in the Netherlands.

An interesting feature of the Dutch factor structures is that, despite the inclusion of many terms related to Intellect, Culture, *and* Openness in the Dutch list, the Intellect and Culture components did not form two separate factors as they did in Goldberg's seven-factor solutions based on the 1,710 list. Moreover, given the presently scant (but suggestive) evidence for a Value factor in the American data, it would be interesting to know whether the Dutch Conservatism factor reflects, in addition to political orientation, the *principled, moral*, and maybe even *religious* characteristics subsumed by that American factor.

GENERAL EVALUATION AND SOME CONCLUSIONS

Taxonomy evaluation criteria

The evaluation and, when warranted, modification of a taxonomic proposal is probably the most important stage in the construction of a taxonomy. Most taxonomy projects have used one or more evaluation criteria that were built directly into the iterative process by which the taxonomy was constructed and refined. Cattell, for example, first aimed for high similarity within and high dissimilarity between his semantic clusters, and then hand-rotated the factors he extracted from peer ratings to a simple-structure position that he found most interpretable. Later he tried to match these factors to others that he derived from analyses of questionnaires and objective-test items. Goldberg tried to maximize the homogeneity of his 42 categories, with respect to both their semantic content and their internal consistency in ratings of oneself and others, in an iterative fashion. Wiggins tried to approximate as closely as possible the angular positions prescribed by his two-dimensional circular model. And finally, Hofstee and Brokken maximized the congruence of their factor structures across self and peer ratings. Unfortunately, with the exception of the last, none of these criteria indicates unambiguously when the iterative process has reached an optimum. That decision thus resides with the taxonomer.

Taxonomies can also be evaluated in terms of their comprehensiveness (or inclusiveness) and the amount of structure they provide. Obviously, alphabetical listings are the least useful; taxonomies that provide a fine-grained structure, such as the circumplex, are of much more utility; hierarchies provide the highest degree of structuring and are most flexibly employed because the availability of various levels of abstraction allows the user to apply the taxonomy to tasks that differ in the degree of descriptive detail and accuracy deemed desirable.

Empirical validation using non-rating correlates

Almost all of the research aimed at evaluating language-based taxonomies has relied on self and peer ratings. Eventually, however, taxonomic structures have to

be evaluated, like other rationally or conceptually constructed measures, with respect to their external validity. In light of their preoccupation with cataloguing and naming, it is not surprising that taxonomers have been said to 'often confuse the invention of a name with the solution of a problem' (Gould, 1981, p. 158). In other words, the existence of a trait name in a natural language does not guarantee that the name refers to something real (Block, Weiss and Thorne, 1979). Nor, for that matter, does the construction of a category formed of semantically similar trait names, or the identification of a factor that summarizes the covariation among ratings of individuals on such trait names. Rather, such categories and such factors suggest the hypothesis that the people in this language community have noticed in themselves or in others a characteristic that is salient in their daily transactions and relates, in systematic ways, to social outcomes they regard as important. However, the systematic test of these hypotheses has barely begun.

Evidence for the external validity of lexically-derived categories and factors is, at present, limited to a few broad concepts, such as the Big Five. For example, peer ratings of Conscientiousness have been shown to predict various indicators of academic achievement (see Wiggins, 1973). Longitudinal research, such as Conley's (1985a, b) analyses based on the Kelly Longitudinal Study, demonstrate that dimensions closely resembling the Big Five, when rated by spouse and self some 45 years earlier, show remarkable stability over time and predict important life outcomes, such as alcoholism, emotional disorders, life satisfaction, and divorce. Finally, McCrae and Costa (1985, 1987) have embarked on an extensive programme of research specifically designed to explore and validate the Big Five dimensions in a large sample of normal adults who are being followed longitudinally.

However, broad-based attempts to study the external validity of the taxonomic structures reviewed here might be premature. The available taxonomies differ in numerous ways, such as in the instances included in the taxonomy, the kind of structure chosen to represent the domain under investigation, and the kinds of data from which they have been derived. These differences partly reflect the varying definitions of personality and research objectives of the taxonomers. However, there are methodological issues as well. For example, Norman, Goldberg, and Wiggins all started from essentially the same set of trait terms yet their taxonomies are remarkably different, both in their general structure and in the specific details. Moreover, there has been no attempt to relate these various taxonomies to each other and integrate them, nor is it clear how that could be done. Even at the level of the broadest dimensions or factors, disagreements continue. Systematic progress and convergence are unlikely to occur unless some generally accepted procedures for the construction of taxonomies can be established, at least for those steps that follow the definition of the universe of personality attributes to be included.

One way to overcome some of the subjectivity and heterogeneity in taxonomic methods and procedures is to ensure, at every step, that the postulated structure is generalizable across data sources, judges and, ideally, cultures and languages. Demonstrations of cross-language generalizability, although central from the vantage point of the lexical approach, have been limited to concepts at the highest level of abstraction. The comprehensive analyses in Dutch have provided so far the strongest cross-language evidence for the Big Five. Results from a study of English-German bilinguals indicate that the Big Five form internally consistent and

relatively independent dimensions in German as well (John *et al.*, 1984). Finally, factor analyses of translations of Norman's (1963) 20 scales have replicated the Big Five in Japanese (Bond, Nakazato and Shiraishi, 1979), although it should be noted that the English-Japanese convergence was low for Factor V.

The construction of a taxonomy of personality-descriptive terms that is generally accepted in the field will require a substantial effort by personality psychologists working in different languages and cultures. But it would be worth the effort. Such a taxonomy could provide a common framework for personality research guided by different theoretical orientations and bring some order to the myriad of individual personality-variables that psychologists have studied over the years.

ACKNOWLEDGEMENTS

This research was supported, in part, by Grant OZ-2780 from the University of Bielefeld, and by Grant MH-39077 from the National Institute of Mental Health (U.S.A.). Moreover, the University of Bielefeld provided travel support for A. Angleitner's stay at the Institute for the Measurement of Personality (Director: Lewis R. Goldberg) and the University of Oregon, Eugene, during the summer of 1983. This trip was also supported by the German Research Foundation (DFG, AN106/5-1). We would like to thank Jack Block, Peter Borkenau, William F. Chaplin, Lewis R. Goldberg, Sarah E. Hampson, Willem K. B. Hofstee, Warren Norman, Rainer Riemann, and Robert A. Wicklund for their helpful comments on an earlier draft of this paper.

REFERENCES

- Allport, G. W. (1937). *Personality: A Psychological Interpretation*, Holt, New York.
- Allport, G. W. (1961). *Pattern and Growth in Personality*, Holt, Rinehart & Winston, London.
- Allport, G. W. and Odbert, H. S. (1936). 'Trait-names: A psycho-lexical study', *Psychological Monographs*, **47**: No. 211.
- Anderson, N. H. (1968). 'Likeableness ratings of 555 personality-trait words', *Journal of Personality and Social Psychology*, **9**: 272-279.
- Angleitner, A., John, O. P. and Loehr, F.-J. (1986). 'It's *what* you ask and *how* you ask it: An itemmetric analysis of personality questionnaires'. In: Angleitner, A. and Wiggins, J. S. (Eds), *Personality Assessment via Questionnaires: Current Issues in Theory and Measurement*, pp. 61-108, Springer-Verlag, Berlin.
- Angleitner, A., Ostendorf, F. and John, O. P. (in preparation). *Towards a Taxonomy of German Personality-descriptors*.
- Baumgarten, F. (1933). 'Die Charaktereigenschaften'. [The character traits]. In: *Beitraege zur Charakter- und Persoenlichkeitsforschung* (Whole No. 1). A. Francke, Bern, Switzerland.
- Becker, W. C. (1960). 'The matching of behavior rating and questionnaire personality factors', *Psychological Bulletin*, **57**: 201-212.
- Block, J., Weiss, D. S. and Thorne, A. (1979). 'How relevant is a semantic similarity interpretation of personality ratings?', *Journal of Personality and Social Psychology*, **37**: 1055-1074.

- Bond, M. H., Nakazato, H. and Shiraishi, D. (1975). 'Universality and distinctiveness in dimensions of Japanese person perception', *Journal of Cross-Cultural Psychology*, **6**: 346-355.
- Borgatta, E. F. (1964). 'The structure of personality characteristics', *Behavioral Science*, **9**: 8-17.
- Brokken, F. B. (1978). *The Language of Personality*, Krips, Meppel.
- Bromley, D. B. (1977). *Personality Description in Ordinary Language*, Wiley, London.
- Burt, C. L. (1937). 'The analysis of temperament', *British Journal of Medical Psychology*, **17**: 158-188.
- Buss, D. M. and Craik, K. H. (1980). 'The act frequency concept of dispositions: Dominance and prototypically dominant acts', *Journal of Personality*, **48**: 379-392.
- Buss, D. M. and Craik, K. H. (1981). 'The act frequency analysis of interpersonal dispositions', *Journal of Personality*, **49**: 14-192.
- Buss, D. M. and Craik, K. H. (1983). 'The act frequency approach to personality', *Psychological Review*, **90**: 105-126.
- Cattell, R. B. (1933). 'Temperament tests: I. Temperament', *British Journal of Psychology*, **23**: 308-329.
- Cattell, R. B. (1934). 'Temperament test: II. Tests', *British Journal of Psychology*, **24**: 20-49.
- Cattell, R. B. (1936). *A Guide to Mental Testing*, University of London Press, London.
- Cattell, R. B. (1943a). 'The description of personality. II. Basic traits resolved into clusters', *Journal of Abnormal and Social Psychology*, **38**: 476-507.
- Cattell, R. B. (1943b). 'The description of personality. I. Foundations of trait measurement', *Psychological Review*, **50**: 559-594.
- Cattell, R. B. (1944). 'Interpretation of the twelve primary personality factors', *Character and Personality*, **13**: 55-90.
- Cattell, R. B. (1945a). 'The description of personality: Principles and findings in a factor analysis', *American Journal of Psychology*, **58**: 69-90.
- Cattell, R. B. (1945b). 'The principal trait clusters for describing personality', *Psychological Bulletin*, **42**: 129-161.
- Cattell, R. B. (1946). *Description and Measurement of Personality*, World Book, New York.
- Cattell, R. B. (1947). 'Confirmation and clarification of primary personality factors', *Psychometrika*, **12**: 197-220.
- Cattell, R. B. (1948). 'The primary personality factors in women compared with those in men', *British Journal of Psychology*, **1**: 114-130.
- Cattell, R. B. (1957). *Personality and Motivation Structure and Measurement*, World Book, New York.
- Cattell, R. B. (1979). *Personality and Learning Theory. Volume I. The Structure of Personality in its Environment*, Springer, New York.
- Chaplin, W. F., John, O. P. and Goldberg, L. R. (1988). 'Conceptions of states and traits: Dimensional attributes with ideals as prototypes', *Journal of Personality and Social Psychology*, in press.
- Conley, J. J. (1985a). 'Longitudinal stability of personality traits: A multitrait-multimethod-multioccasion analysis', *Journal of Personality and Social Psychology*, **49**: 1266-1282.
- Conley, J. J. (1985b). 'A personality theory of adulthood and aging'. In: Hogan, R. and Jones, W. H. (Eds), *Perspectives in Personality*, Vol. 1, pp. 81-115, JAI Press, Greenwich.
- Digman, J. M. (1963). 'Principal dimensions of child personality as seen in teachers' judgments', *Child Development*, **34**: 43-60.
- Digman, J. M. (1972). 'The structure of child personality as seen in behavior ratings'. In: Dreger, R. M. (Ed.), *Multivariate Personality Research*, pp. 587-611, Claitor's, Baton Rouge.
- Digman, J. M. and Inouye, J. (1986). 'Further specification of the five robust factors of personality', *Journal of Personality and Social Psychology*, **50**: 116-123.
- Digman, J. M. and Takemoto-Chock, N. K. (1981). 'Factors in the natural language of personality: Re-analysis and comparison of six major studies', *Multivariate Behavioral Research*, **16**: 149-170.

- Fenigstein, A., Scheier, M. and Buss, A. H. (1975). 'Public and private self-consciousness: Assessment and theory', *Journal of Consulting and Clinical Psychology*, **43**: 522-527.
- Fiske, D.W. (1949). 'Consistency of the factorial structures of personality ratings from different sources', *Journal of Abnormal and Social Psychology*, **44**: 329-344.
- Foa, U. G. and Foa, E. B. (1974). *Societal Structures of the Mind*, Charles C. Thomas, Springfield.
- Galton, S. F. (1884). 'Measurement of character', *Fortnightly Review*, **42**.
- Goldberg, L. R. (1980). 'Some ruminations about the structure of individual differences: Developing a common lexicon for the major characteristics of human personality', Invited paper, Convention of the Western Psychological Association, Honolulu, Hawaii.
- Goldberg, L. R. (1981). 'Language and individual differences: The search for universals in personality lexicons'. In: Wheeler, L. (Ed.), *Review of Personality and Social Psychology*, Vol. 2, pp. 141-165, Sage, Beverly Hills.
- Goldberg, L. R. (1982). 'From Ace to Zombie: Some explorations in the language of personality'. In: Spielberger, C. D. and Butcher, J. N. (Eds), *Advances in Personality Assessment*, Vol. 1, pp. 203-234, Erlbaum, Hillsdale.
- Gough, H. G. and Heilbrun, A. B. Jr. (1965). *The Adjective Check List Manual*, Consulting Psychologists Press, Palo Alto, California.
- Gould, S. J. (1981). *The Mismeasure of Man*, Norton, New York.
- Guilford, J. P. and Zimmerman, W. S. (1956). 'Fourteen dimensions of temperament', *Psychological Monographs*, **70**: (10, Whole No. 417).
- Hampshire, S. (1953). 'Dispositions', *Analysis*, **14**: 5-11.
- Hampson, S. E. (1982). 'Person memory: A semantic category model of personality traits', *British Journal of Psychology*, **73**: 1-11.
- Hampson, S. E., Goldberg, L. R. and John, O. P. (1987). 'Category-breadth and social-desirability values for 573 personality terms', *European Journal of Personality*, **1**: 241-258.
- Hampson, S. E., John, O. P. and Goldberg, L. R. (1986). 'Category breadth and hierarchical structure in personality: Studies of asymmetries in judgments of trait implications', *Journal of Personality and Social Psychology*, **51**: 37-54.
- Hofstee, W. K. B. (1976). *Dutch traits: The first stages of the Groningen taxonomy study of personality descriptive adjectives*, University of Groningen, The Netherlands.
- Hofstee, W. K. B. (1977). 'Preliminary steps in structuring the set of 1,204 Dutch personality-descriptive adjectives', Heymans Bulletins, HB-77-302 IN, University of Groningen, The Netherlands.
- Hofstee, W. K. B., Brokken, F. B. and Land, H. (1981). 'Constructie van een Standaard-Persoonlijkheids-Eigenschappenlijst (S.P.E.L.)'. [Construction of a Standard Personality Trait List]. *Nederlands Tijdschrift voor de Psychologie*, **36**: 443-452.
- Howarth, E. (1976). 'Were Cattell's "personality sphere" factors correctly identified in the first instance?' *British Journal of Psychology*, **67**: 213-230.
- John, O. P. (1986). 'How shall a trait be called? A feature analysis of altruism'. In: Angleitner, A., Furnham, A. and Van Heck, G. (Eds), *Personality Psychology in Europe: Current Trends and Controversies*, Vol. 2, pp. 117-140. Swets North America, Berwyn.
- John, O. P., Goldberg, L. R. and Angleitner, A. (1984). 'Better than the alphabet: Taxonomies of personality-descriptive terms in English, Dutch, and German'. In: Bonarius, H. C. J., Van Heck, G. L.M. and Smid, N. G. (Eds), *Personality Psychology in Europe: Theoretical and Empirical Developments*, Vol. 1, pp. 83-100, Swets & Zeitlinger, Lisse.
- Klages, L. (1926). *The Science of Character* (Translated 1932), Allen & Unwin, London.
- McCrae, R. R. and Costa, P. T. (1985). 'Updating Norman's adequate taxonomy: Intelligence and personality dimensions in natural language and in questionnaires', *Journal of Personality and Social Psychology*, **49**: 710-721.
- McCrae, R. R. and 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.
- McDougall, W. (1923). *An Outline of Psychology*, Scribners, New York.
- Norman, W. T. (1963). 'Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings', *Journal of Abnormal and Social Psychology*, **66**: 574-583.

- Norman, T. (1967). '2,800 personality trait descriptors: Normative operating characteristics for a university population', Department of Psychology, University of Michigan.
- Nowakowska, M. (1973). 'The limitations of the factor-analytic approach to psychology with special application to Cattell's research strategy', *Theory and Decision*, **4**: 109-139.
- Peabody, D. (1967). 'Trait inferences: Evaluative and descriptive aspects', *Journal of Personality and Social Psychology*, **7**: (4, Whole No. 644).
- Peabody, D. (1968). 'Group judgments in the Philippines: Evaluative and descriptive aspects', *Journal of Personality and Social Psychology*, **10**: 290-300.
- Peabody, D. (1970). 'Evaluative and descriptive aspects in personality perception: A reappraisal', *Journal of Personality and Social Psychology*, **16**: 639-646.
- Peabody, D. (1987). 'Selecting representative trait adjectives', *Journal of Personality and Social Psychology*, **52**: 59-71.
- Peabody, D. and Goldberg, L. R. (1987). *Variance and Invariance in Personality Structures: Determinants of Factors Derived from Trait Adjectives*, Oregon Research Institute, Eugene.
- Sells, S. B. (1959). 'Structured measurement of personality and motivation: A review of contributions of Raymond B. Cattell', *Journal of Clinical Psychology*, **15**: 3-21.
- Sheldon, W. H. and Stevens, S. S. (1942). *The Varieties of Temperament: A Psychology of Constitutional Differences*, Harper, New York.
- Simpson, G. G. (1961). *Principles of Animal Taxonomy*, Columbia University Press, New York.
- Snyder, M. (1974). 'Self-monitoring of expressive behavior', *Journal of Personality and Social Psychology*, **30**: 526-537.
- Spearman, C. (1927). *Abilities of Man*, Macmillan, London.
- Spoehr, H. D. (1936). 'Faculties versus traits: The solution of Franz Joseph Gall', *Character and Personality*, **4**: 216-231.
- Spranger, E. (1914). *Lebensformen. Geisteswissenschaftliche Psychologie und Ethik der Persoenlichkeit* [Types of men]. Halle (a. d. Saale), Niemeyer, Germany.
- Tupes, E. C. and Christal, R. C. (1961). 'Recurrent personality factors based on trait ratings', Technical Report, USAF, Lackland Air Force Base, TX.
- Webster's New International Dictionary* (1925). 2nd unabridged edition, Merriam, Springfield.
- Webster's New International Dictionary* (1961). 3rd unabridged edition, Merriam, Springfield.
- Wiggins, J. S. (1973). *Personality and Prediction: Principles of Personality Assessment*, Addison-Wesley, Reading.
- Wiggins, J. S. (1974). 'In defense of traits', Invited paper, Ninth Annual Symposium on the Use of MMPI, Los Angeles, CA.
- Wiggins, J. S. (1979). 'A psychological taxonomy of trait-descriptive terms: The interpersonal domain', *Journal of Personality and Social Psychology*, **37**: 395-412.
- Wiggins, J. S. (1980). 'Circumplex models of interpersonal behavior'. In: Wheeler, L. (Ed.), *Review of Personality and Social Psychology*, Vol. 1, pp. 265-294, Sage, Beverly Hills.
- Wiggins, J. S. and Broughton, R. (1985). 'The interpersonal circle: A structural model for the integration of personality research'. In: Hogan, R. and Jones, W. H. (Eds), *Perspectives in Personality*, Vol. 1, pp. 1-47, JAI Press, Greenwich.
- Wittgenstein, L. (1953). *Philosophical Investigations*, Macmillan, New York.

RÉSUMÉ

Nous décrivons dans cette bibliographie des taxinomies de termes descriptifs de la personnalité. Nous avons identifié cinq problèmes de l'établissement de taxinomies et nous évaluons les avantages, ainsi que les inconvénients de l'approche lexicographique. Notre bibliographie commence par la liste des traits caractéristiques d'Allport et Odbert, examine les procédures qui ont mené aux facteurs de la personnalité de Cattell, et résume les travaux contemporains anglais et néerlandais. Les chercheurs qui construisent ces taxinomies considèrent, tout comme les profanes, les *traits stables* comme les concepts de la personnalité

les plus fondamentaux. Ce domaine, à son plus haut niveau d'abstraction, est souvent représenté par cinq dimensions relatives au *Pouvoir* (ou Surgency), à *l'Amour* (Amabilité), au *Travail* (Conscience), à *l'Affect* (Stabilité émotionnelle) et à *l'Intellect* (Culture). Ces dimensions toutefois, étant donné le fait que l'on débat encore de leur nombre et de leur interprétations, ne doivent pas être considérées comme définitives mais comme un cadre heuristique utile. Notre synopsis suggère que, lorsque l'on interprète la cinquième dimension comme Intellect, Culture, Valeurs et même Autonomie-Conformité entrent en ligne de compte comme dimensions additionnelles. Quelques chercheurs tels Goldberg et Wiggins ont construit des catégories plus étroites afin de fournir une reproduction plus différenciée des termes descriptif. Il faudrait encore développer des taxinomies d'états, activités et rôles sociaux et effets; santé, forme physique et attirance physique devraient également être pris en considération. Une taxinomie détaillée pourvoit d'un cadre commun les expériences réalisées à partir d'orientations théoriques différentes dans la psychologie de la personnalité et faciliterait la sélection systématique des variables pour la recherche. Enfin, la valeur d'une taxinomie dépend de son succès à prédire les événements importants de la vie.

ZUSAMMENFASSUNG

Die Arbeit gibt einen Überblick über den Forschungsstand und die Entwicklung von Taxonomien persönlichkeitsbeschreibender Begriffe. Im Hinblick auf die Konstruktion einer Taxonomie werden fünf zentrale Probleme herausgestellt und die Vor- und Nachteile des lexikalischen Ansatzes erörtert. Ausgehend von der Darstellung der Arbeiten Allport's und Odbert's schildern wir die einzelnen Forschungsschritte, die Cattell zur Annahme seiner Persönlichkeitsfaktoren geführt haben und fassen den gegenwärtigen Forschungsstand amerikanischer und holländischer Taxonomien zusammen. *Stabile Eigenschaften* werden sowohl von Forschern, als auch von Laien als die grundlegendsten Persönlichkeitskonzepte angesehen. Die gesamte Klasse der stabilen Eigenschaftsbegriffe wird auf einem sehr generellen Abstraktionsniveau häufig durch fünf Dimensionen repräsentiert, die mit den Etiketten *Macht* (oder Surgency), *Liebe* (oder Agreeableness), *Arbeitsverhalten* (oder Conscientiousness), *Affekt* (oder Emotional Stability) und *Intellect* (oder Culture) bezeichnet werden. Da die Interpretation und Anzahl dieser Dimensionen allerdings noch umstritten ist, sind sie als vorläufiger Bezugsrahmen anzusehen. Wird die fünfte Dimension als Intellect interpretiert, so kommen Kultiviertheit, Moralische Werthaltung und vielleicht Autonomie-Konformität als weitere Dimensionen in Betracht. Einige Forscher, wie z.B. Goldberg und Wiggins, haben spezifischere Kategorien entwickelt, um eine differenziertere Taxonomie eigenschaftsbeschreibender Begriffe zu erreichen. Für persönlichkeitsbeschreibende Begriffe, die kurzfristige Zustände, Verhaltensweisen, und soziale Rollen und Effekte beschreiben, müssen noch Taxonomien entwickelt werden. Ebenfalls sollten Begriffe, die sich auf die Gesundheit, Fitneß und die körperliche Attraktivität von Personen beziehen, berücksichtigt werden. Eine umfassende Taxonomie würde einen gemeinsamen Bezugsrahmen für verschiedene theoretische Orientierungen in der Persönlichkeitspsychologie bereitstellen und die systematische Auswahl von Variablen in der Forschung erleichtern. Letztlich hängt der Wert einer jeden Taxonomie davon ab, ob durch sie wichtige Lebensereignisse vorausgesagt werden können.