The Spanish Journal of Psychology
2007, Vol. 10, No. 1, 68-83

Copyright 2007 by The Spanish Journal of Psychology
ISSN 1138-7416

Constructing Personality Maps, Mapping Personality Constructs: Multidimensional Scaling Recovers the Big Five Factors from Internal and External Structure

David Bimler and John Kirkland
Massey University, New Zealand

This report examines the structure of similarities underlying the lexicon of personality-trait description, when "similarity" is defined and measured in terms of (a) semantic judgment and (b) covariance in actual use. A lexicon of 60 trait adjectives was examined, using several procedures for collecting semantic judgments. Similarity data of both kinds were analyzed with multidimensional scaling (MDS) to provide a parsimonious representation of underlying structure. The convergence between semantic judgments and covariance within trait-attribution data was substantial; both kinds of data evinced the same structure when collected for subsets of adjectives. Canonical correlation was employed to find the number of dimensions shared across MDS solutions. Interpretation of the results was facilitated by individual-differences MDS, which can select an optimal set of underlying dimensions, and at the same time accommodate the differences between data sets that arise when data-collection procedures differ in the relative emphasis they place upon those dimensions. We interpret the small number and shared nature of the dimensions by arguing that the lexicon's structure relates to trait perception rather than personality structure per se, even when probed with trait-attribution covariance.
Keywords: Big Five, trait perception, internal structure, multidimensional scaling

Este trabajo examina la estructura de las similitudes subyacentes al léxico de la descripción de los rasgos de personalidad, cuando "similitud" se define y se mide en términos de: (a) juicio semántico y (b) covarianza en el uso actual. Se examinó un léxico de 60 adjetivos de rasgos, empleando varios procedimientos para recoger juicios semánticos. Los datos de similitud de ambos tipos se analizaron con escalonamiento multidimensional (EMD) para obtener una representación parsimoniosa de la estructura subyacente. La convergencia entre los juicios semánticos y la covarianza rasgo-datos atribucionales era sustancial; ambos tipos de datos mostraban la misma estructura cuando se recogían para subconjuntos de adjetivos. Se empleó la correlación canónica para encontrar el número de dimensiones compartidas por las soluciones EMD. La EMD de diferencias individuales facilitó la interpretación de los resultados porque puede seleccionar un conjunto óptimo de dimensiones subyacentes y, al mismo tiempo, adaptar las diferencias entre los conjuntos de datos que emergen cuando los procedimientos de recogida de datos difieren con respecto al énfasis relativo que se concede a dichas dimensiones. Nosotros interpretamos el pequeño número y la naturaleza compartida de las dimensiones arguyendo que la estructura del léxico se relaciona más con la percepción de los rasgos que con la estructure de la personalidad en sí, incluso cuando se analiza mediante la covarianza rasgo-atribución.
Palabras clave: Big Five, la percepción de rasgos, la estructura interna, escalonamiento multidimensional

---

Acknowledgements: This research would have been impossible without the informants, including students who found the task of arranging words into groups more congenial than formal lessons; the school principals who gave them the choice; or the assistants with the work of data collection. Reading the handwriting of informants in many cases called for advanced cryptanalytic skills, so a special thanks is owed to Robyn Surville and Lynne McMorran who wrestled with loosely-structured data-collection forms as they entered GOPA-sorting data. We are grateful for constructive comments from the anonymous reviewers of a previous version of this report.

Address correspondence concerning this article to John Kirkland, School of Arts, Development and Health Education, College of Education, Massey University, Private Bag 11-222, Palmerston North, New Zealand. E-mail: j.kirkland@massey.ac.nz

Over the last two or three decades, the relationships and connections among descriptors of personality traits have become a major focus of empirical research, as well as the subject of interest from the theoretical and applied communities. To put it in very broad terms, the hope is that a relatively simple set of variables might hold sway above the plethora of trait-descriptive words, phrases, and sentences. Reduction to these variables would provide the language for an adequate description of an individual's personality structure, accommodate much of the variation between individual structures, and be universal, in the sense that the variables could ideally be translated across languages and cultures without loss of validity.

The structure underlying an inventory of descriptors can be made explicit in a variety of ways. Following Wiggins (1973), it is convenient to distinguish between *internal* and *external structure*. The former is evinced when informants treat descriptors as free-floating abstractions, and make judgments about *similarities of meaning* among them rather than apply them to specific personalities (D'Andrade, 1965). In a related methodology, observers estimate the probability with which the presence of trait A predicts trait B (Hakel, 1969).

Conversely, external structure applies to an empirical pattern of correlations, among numerical values obtained by asking observers to rate the descriptors on how well they describe target individuals (with self-descriptions as a special case). Such correlations are typically approached with factor analysis (FA). The results provide theoretical clues to any structure that may underlie the inventory, and also a practical guide to summarizing a large body of ratings in terms of a smaller number of aggregate scales.

Numerous studies of external structure have crystallized around a specific framework of personality description: the 'Big-Five' factor structure (see Carroll, 2002, for a critical review). Different names have been used over the years to characterize the factors of the Big Five. Traditional names are as follows, with shorthand abbreviations, and illustrative bipolar oppositions within parentheses: Extraversion, Ex (*extrovert-introvert*); Agreeableness, A (*warm-cold*); Conscientiousness, C (*organized-disorganized*); Emotional Stability, ES (*calm-neurotic*); and Intellect or Openness to Experience, OE (*intelligent-unintelligent*). This is related to (but not necessarily identical to) the "five-factor model," or FFM. While researchers may agree about the need for a five-factor solution, they disagree about the exact meaning of each factor, or give them brief and imprecise labels. Close scrutiny of the descriptions in the "small print" reveals that there is more than one five-factor model.

Crucially, internal data deliver convergent results. For instance, Sneed, McCrae, and Funder (1998) examined the semantic structure among 30 trait adjectives. In their Study 1, informants scored the items for "indicativeness" of the FFM factors, that is, how closely each one approached the core meanings of one pole of each factor in turn. In their Study 3, the items were simply sorted into groups according to similarity. The results cannot confirm the paramount status of any particular factors, because factor poles were provided in both cases (implicitly in Study 3, by the tight clustering of the items). However, the results were compatible with the FFM structure determined empirically from trait-attribution data.

The research reported here is intended to explore this external / internal convergence in depth, exploring (among other things) how many dimensions the two forms of structure have in common, and how those dimensions can best be interpreted. The internal structure of 60 trait adjectives (i.e., their meanings) was compared against their Big-Five properties, in the form of their factor loadings in published FA results. To this end, we elicited ratings of the indicativeness of the items, as examples of the poles of all five factors (as in Sneed et al., 1998). For an unconstrained examination of the items' semantic structure, we elicited judgments of the similarities and dissimilarities among them. Informants employed a battery of similarity-sorting procedures, and a triadic method, as a precaution against the possible inadequacy of any single procedure.

Note that similarity judgments require the simultaneous consideration of at least two items. When participants sort items into groups, all items are presented simultaneously, so that decisions are made in the context of the entire item set (e.g., Church & Katigbak, 1998; Sneed et al., 1998). For consistency, we probed the external structure of the 60 descriptors by using a Q-sorting or ranking procedure to elicit peer descriptions; this technique again sets the items in context by presenting them simultaneously while they are assessed. This contrasts with the more usual Likert-style peer-rating task where each word (or phrase, or sentence) is rated in isolation from other items in the inventory. By aligning procedures, we remove one potential source of discrepancy between trait attributions and similarity judgments.

Similarity data lend themselves to analysis with multidimensional scaling (MDS). This operates on the data to construct a geometrical representation of the items as points in a "stimulus space." Two items can differ from each other in various ways, with each mode of difference contributing to the total dissimilarity between them, and each dimension in a MDS solution corresponds to a separate mode. To facilitate comparisons and eliminate "method variance," we also applied MDS to trait attribution data, instead of the more usual FA. There is much in common between the mathematical models used by MDS and FA, and between the outcomes they deliver, although they diverge in some respects (Davison & Skay, 1991).

Since the dimensionality of "trait space" is of particular interest, spatial configurations (MDS solutions) from the various forms of data were compared using canonical correlation, to check how many dimensions they shared in common. It is plausible that the different tasks might vary

by being more or less sensitive to the underlying dimensions, that is, by placing varying weights (importance, salience) upon each source of dissimilarity or form of inter-item difference. This possibility was accommodated by using an appropriate model of *individual-differences* MDS. Such models facilitate the interpretation of the results as they break the rotational indeterminacy typically found in MDS and FA solutions.

Also of interest is whether the solutions reveal a simplex or circumplex structure, in which items are distributed continuously. Johnson and Ostendorf (1993) found a geometrical circumplex structure necessary to clarify and to accommodate the discrepancies between different *forms* of five-factor model. Circumplex models have often been put forward for the inter-personal subspace of the personality-trait domain, to complement rather than supplant the Big-Five model (e.g., Hofstee, de Raad, & Goldberg, 1992). It has been argued that MDS is preferable in such situations, because principal-components analysis (the simplest form of FA: Harris, 1975) can lead to a proliferation of factors that obscure the underlying simplicity (Guttman, 1966; Maraun, 1997).

Several explanations have been advanced for the convergence of internal and external forms of judgment; we revisit this point in the Discussion. For now, suffice to say that the dichotomy is not absolute. Although similarity judgments are clearly derived from the "meanings" of the terms, those meanings are also involved in the task of self- or other-rating. While describing a peer, if observers have incomplete information, they may have to extrapolate or impute their ratings according to an implicit model of personality, an IPT (Passini & Norman, 1966). For instance, their access to the target may be limited to a brief encounter. Their responses reveal the cognitive schemata they deploy to organize and draw inferences from their experiences of behavior. The internal relative relationships among trait descriptors will generally have some influence on how they are applied to a target. If a peer-description uses the terms in a way that departs from their meanings, it is simply incoherent.

Qualities like "altruism" or "excitability" are explanations for patterns of behavior— patterns with other possible explanations—rather than tangible, concrete entities. We argue that one does not learn the nouns and noun-phrases of traits by observing an entity and associating it with the corresponding term. Instead, as we learn these terms in the course of language acquisition, the knowledge we gain about each one consists of its position relative to others in the same domain—its connections to near-synonyms or antonyms. If we conform to that pattern of relationships, when we go on to use the terms to characterize the behavior of others or ourselves, then our listeners will grant that we speak the language properly. Conversely, if we apply the terms in a way inconsistent with the usage of others, they will not conclude that our experiences of (e.g.) "altruism" were atypical, but rather that our grasp of the language is inadequate.

Method

Stimuli

Goldberg (1992, Table 3) lists 10 single-word personality-trait markers for each pole of the Big-Five factors (with one exception: only six words were linked sufficiently tightly with ES+ to be markers). The six items with the highest absolute loadings on each factor pole were used as items here, 60 words in all (*imperturbable* in the ES+ pole was replaced with its more familiar near-synonym *calm*). These were printed in 14-pt. Times Roman Bold on slips of card measuring 60 × 30 mm, along with a three-digit tag for purposes of data entry.

These items were selected as representative of the trait lexicon (Goldberg, 1992), but they are not spread evenly through it. Thus, it is worth noting that any clustering found among the items will partly be the result of this process of selection, which removed any traits that loaded moderately on two or more of the Big-Five factors. This should not affect the dimensionality of the "trait space" required to accommodate the items.

Participants

Groups of adults (age 25-45) were recruited opportunistically: 20 as Sample 1, 15 as Sample 4, and 24 as Sample 5. Groups of school students (age 13-15) were recruited from local secondary schools, with the cooperation of the teaching staff and the consent of the students themselves. Twenty-three took part as Sample 2, 47 as Sample 3, 19 and 17 as Sample 6, 24 as Sample 7, and 39 as Sample 8.

Some of the student participants were unfamiliar with some of the items. Such words were given ad hoc definitions, or else they were omitted. Graziano, Jensen-Campbell, Steele, and Hair (1998) found that although such omissions are not distributed randomly through the lexicon, they do not affect its derived factorial structure.

Procedures

Sample 1. Participants followed a two-phase procedure that we have called GA-sorting, for Grouping and Addition. The Grouping phase consisted of arranging items into groups on the basis of perceived similarity, using participants' own criteria as to what constitutes "similarity." The number of items per group was unrestricted (single-item groups were permitted); a range of 10 to 20 was suggested as a suitable number of groups. In the Addition phase, they merged the groups in a series of steps (selecting the two most similar groups at each step and combining them into one), until further integration was not possible (Sherman, 1972). The membership of each group was recorded after the initial grouping and after each subsequent merging step.

Sample 2. A more elaborate GPA-sorting task (Grouping, Partition, Addition) was followed with this sample, to provide an insight into the arrangement of items within the groups. After the original Grouping phase, participants were instructed to create and record a finer subdivision (i.e., a partition with more groups). They did this by inspecting each of the groups they had initially created, deciding whether the items comprising it were homogeneous in meaning, and if not, how it could be broken into subgroups. After recording subgroups from this Partition phase, if any, they continued as before with the successive merging (Addition) phase.

We were concerned that the procedure lent itself to stereotyped results. If the 60 items fall into 10 clusters of six relatively synonymous factor markers (as predicted by the Big-Five model), then the sorting task has an easily-recognizable "right answer" (as in Study 3 of Sneed et al., 1998). To ensure that participants each made their similarity judgments in a different, incomplete context, they sorted random subsets of items. This was done by instructing each participant to shuffle the item cards into two subsets of 30, which were sorted separately.

Sample 3. The information from GPA-sorting primarily describes the small similarities between similar items. A GOPA-sorting task complemented this with information about large dissimilarities, obtained in an "Opposite-sorting" phase (O-phase), after the G-sorting phase. Sorters were instructed to choose the two groups out of those they had constructed which provided the strongest contrast or antinomy, that is, the pair which were most opposite in meaning. Putting those groups aside, they could choose a second pair of opposites, up to three times. They then restored the groups in preparation for Partition-sorting. All 60 items were used, again raising the possibility that stereotyped "right answers" might influence participants' decisions, but this time exploiting it as a source of additional data. A secondary purpose of this procedure was to discourage the merging of diametrically-opposite item-groups in the A-phase. Otherwise, such exact antonyms can be seen as *similar*, because they share a common theme.

Sample 4. Each participant ranked the items along five semantic gradients, intended to reproduce the Big-Five factors. Participants received five copies of a template showing empty cells ready for item ID tags, arranged in nine columns with a quasi-normal distribution (columns contained 3, 4, 7, 10, 12, 10, 7, 4, 3 cells). The semantic gradients were defined by printing a cluster of near-synonyms at each extreme of each template. These were additional entries from Goldberg (1992, Table 3), as follows. Although these words have lower loadings on the factors than the 60 items, the overlapping meanings within each cluster should triangulate upon the factor pole.

Ex: [*active, daring, vigorous, unrestrained*] versus [*withdrawn, timid, inactive, unadventurous*].

A: [*pleasant, agreeable, helpful, generous*] versus [*rude, selfish, uncooperative, stingy*].

C: [*careful, hardworking, conscientious, responsible*] versus [*inconsistent, lazy, impractical, sloppy*].

ES: [*placid, peaceful, secure, uninhibited*] versus [*fretful, jealous, nervous, tense*].

OE: [*analytical, innovative, curious, deep*] versus [*unreflective, imperceptive, uninquisitive, shallow*].

Participants were instructed to rank the items five times, each time into the order of one of these polarities, sorting them into nine piles of the sizes provided in the template. They then copied the ID tags into corresponding cells of that template. For the advantages of *ranking* the items in this way (Q-sorting) instead of *rating* them, see Block (1961).

Sample 5. Twenty-four adults arranged the items according to how well they described an acquaintance, in nine ranked piles ranging from "most applicable" to "least applicable," recording the piles in the same templates as Sample 4 (without the synonym poles).

Sample 6. A preliminary examination of the data from Sample 1 yielded an arrangement of the items in a hierarchical tree (a dendrogram). Alternate items were drawn from the tree to yield two subsets of items scattered evenly through the lexicon. Removal of 6 more items reduced the numbers in each subset to 27, with 2 or 3 items representing each cluster of factor-pole markers. Seventeen participants sorted one subset by similarity, following the GPA-sorting task. A further 19 participants did the same for the second subset. The same participants then ranked their respective subsets according to descriptiveness of a peer. They followed a slightly different Q-sorting procedure (the Method of Successive Sorts, or MOSS) to arrange the 27 items in eight ranked piles, without prescribing the numbers in each pile.

Sample 7. A further validation sample used 21 items, selected from the dendrogram to ensure an even sampling of the lexicon. Twenty-four participants Q-sorted these items into seven piles (each of three items) according to descriptiveness of a peer.

Sample 8. A different group of 39 participants followed a triadic procedure with the same 21 items. They shuffled the cards into seven sets of 3, recorded these triads, and indicated the least-similar item—the "odd-one-out"—from each triad. The task can be considered as a limiting case of sorting random subdecks. Each participant repeated the process between one and three times, providing a total of 570 odd-one-out judgments.

Statistical Analysis

For the Q-sorting data from Samples 4-7, we write the k-th Q-sort as a row vector $\mathbf{v}_k$. Within $\mathbf{v}_k$, the i-th entry v_{ki} is the number of the pile in which the i-th item was placed. In Samples 4 and 5, v_{ki} values range from –4 (placed in the left-most pile, at the positive factor pole, or "most applicable") to +4 (right-hand pile, at the negative factor pole or "least applicable"). In Sample 6, they range from 1 to 8. In Sample 7, they range from –3 ("most applicable") to +3 ("least applicable").

Preliminary analysis of the data from Sample 4 found considerable consistency between the 15 Q-sorts for each factor. For each factor f (ranging from Ex to EO), we averaged Sample-4 v_{ki} data over subjects, yielding mean values y_{fi}. This ignores the methodological solecism of averaging ordinal data. For comparison, we took "varimax-rotated factor loadings in [...] descriptions of liked peers" (Goldberg, 1992, Table 3), labeling them as z_{fi}.

Four 60-by-60 matrices of estimated inter-item dissimilarity were derived from separate data sets:

D^R: the semantic Q-sorting of Sample 4;

D^Q: the peer Q-sorting of Sample 5. Individual entries within them are, for instance, d^R_{ij}, the dissimilarity in D^R between the i-th and j-th items. d^R_{ij} and d^Q_{ij} are simply the correlations between values assigned to the i-th and j-th items, v_{ki} and v_{kj}, across the respective Q-sorts.

D^S: derived from the sorting data from the G, P, and A phases of Samples 1, 2, 3, and 6 (no substantial differences emerged between the samples when analyzed separately, so they were pooled); and

D^O: the O-phase data of Sample 3.

Appendices A and B specify the algorithms used to obtain d^S_{ij} and d^O_{ij}.

Hierarchical clustering. The group-mean hierarchical clustering algorithm was used to represent the dissimilarity matrices as tree diagrams or dendrograms.

MDS. The goal of multidimensional scaling is to arrange points in a P-dimensional space so that distances between them reflect the inter-item dissimilarities. We applied MDS to the four similarity matrices separately, using an implementation of Kruskal's nonmetric algorithm. The solution in each case (the points' coordinates) can be written as a 60-by-P matrix, where P ranged from 2 to 6. The poorness-of-fit values (stress) were plotted against P, but in most cases, the curves lacked obvious "elbows" that might indicate an optimum dimensionality, that is, the P for which the solutions reproduced the raw data with both fidelity and parsimony. The presence of "elbows" is a long-standing test for dimensionality, analogous to the scree test in FA, though concerns have been voiced that it can over-estimate the true number of dimensions (Bijmolt & Wedel, 1999).

Five-dimensional solutions (X^S, X^O, X^R, and XQ) were retained to ensure that no structure of importance was lost. The first three matrices were also analyzed together (with $2 \le P \le 6$), combining MDS with the "weighted Euclidean" model of individual differences, to rotate the solution—a "group configuration", X^G_P —to non-arbitrary axes (Kruskal & Wish, 1978).

Canonical correlation. We used canonical correlation (CANCORR) to compare the five-dimensional MDS solutions on a pairwise basis. CANCORR extracts a pair of linear combinations from the two coordinate sets under comparison, such that the correlation between them R_c is maximal (e.g., Harris, 1975). It can extract further pairs of linear combinations of coordinates, providing correlations R_2, R_3, R_4, R_5 (each new combination being orthogonal to those previously extracted from its respective coordinate set). Testing the significance of each successive correlation using Wilks' Λ statistic, the number of significant correlations indicates the number of mutually-recognizable dimensions shared between the two coordinate sets.

The five sets of empirical factor loadings z_f, and the average semantic-gradient values y_f, were also compared against each five-dimensional solution.

Results

For factors Ex through OE, product-moment correlations between respective y_f and z_f were r = 89, .87, .93, .70, and .81, confirming that individuals can indeed recover the five factors from semantic knowledge alone from a pair of poles provided by the experimenter.

In the case of Emotional Stability, many psychometricians would prefer a higher correlation between y_{ES} and z_{ES}, with .70 indicating a less-than-robust recovery. It could be argued that the weakness lies in the adjectives available to epitomize the poles of the factor, and that these did not specify the precise opposition quantified as z_{ES} by Goldberg (1992). We note also that ES is least robust among the Big Five (Peabody & Goldberg, 1989), that is, most affected by changes in procedure. Further, it will be seen below that the ES markers form the least coherent clusters in the hierarchical-clustering and MDS solutions (Figures 1 and 2).

Hierarchical Clustering

Using hierarchical clustering, the GPA-sorting dissimilarities D^S can be represented as a dendrogram (Figure 1). This is a consensus tree, as it were; a compromise over the 149 individual (incomplete) trees provided as sorting data. It contains 10 clusters corresponding to the poles of the factors. At a broad level, items from the positive and negative poles are arranged in two half-trees, joined at the lowest level of similarity, that is, the left-most link in the dendrogram. In other words, the factors are not conceptually orthogonal, and rather than being equidistant from each other, the poles are linked at intermediate levels of similarity. Thus, the A+ and ES+ clusters are closely linked, as are A- and ES– (Agreeableness and Emotional Stability); the distinctions between C+ and OE+, and C– and OE–, are not much stronger (Conscientiousness and Openness). At a lower similarity level, the (C/OE)– group joins with (A/ES)–, leaving the Ex– cluster as the last to join the negative half-tree, whereas the (C/OE)+ grouping is closer to Ex+, leaving (A/ES)+ until last. *Distrustful* from A– is displaced to the C– cluster (it seems that participants understood it as "untrustworthy"). Two ES+ items, *unemotional* and *unexcitable*, have become linked to Ex–.

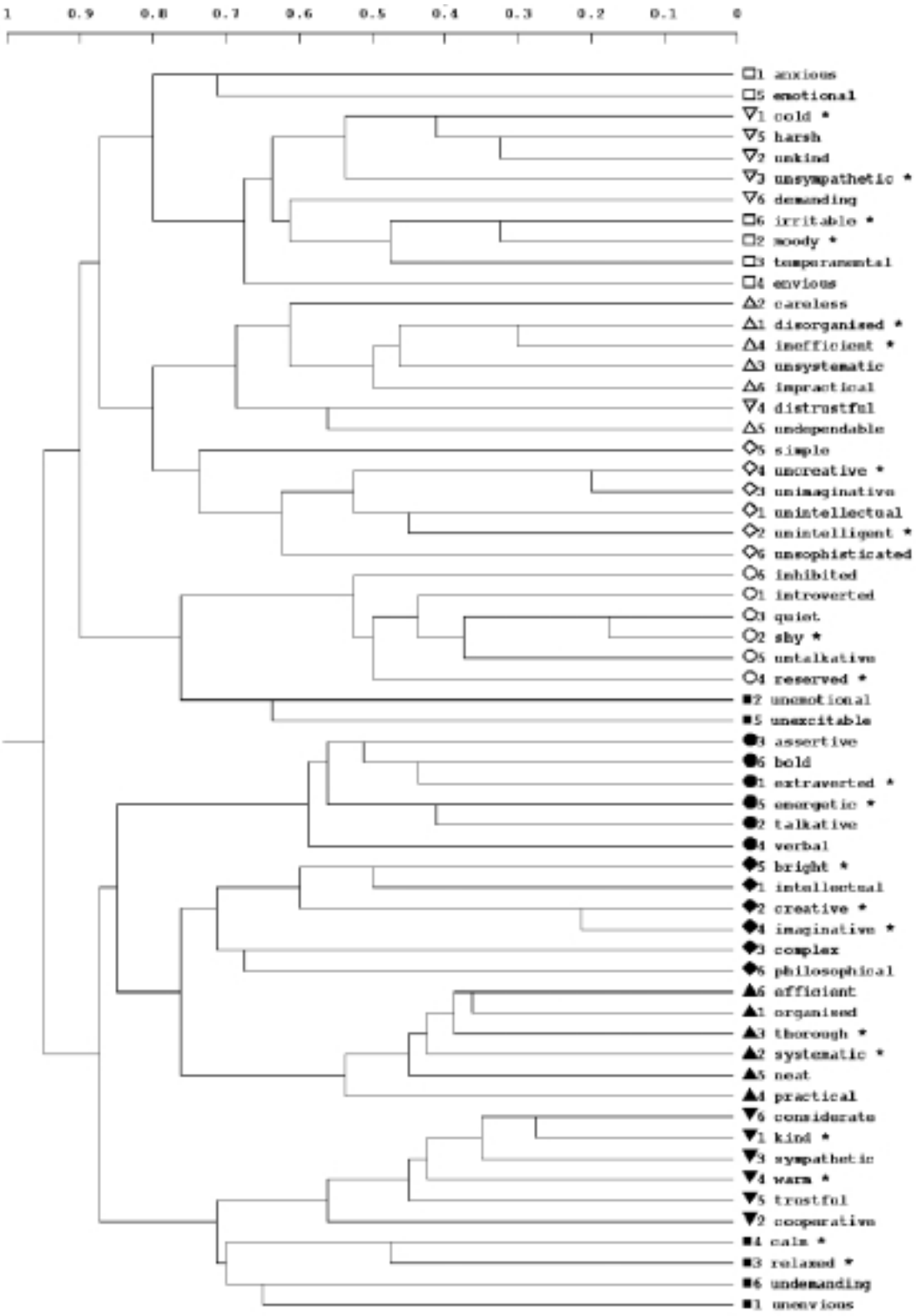

Figure 1. Consensus hierarchical tree summarizing dissimilarity estimates derived from 20 GA-sorts and 129 GPA-sorts. Items are labeled by the Big-5 factor they mark, as follows:

Positive pole of Factor **Ex**: ● Factor **A**: ▼ Factor **C**: ▲ Factor **ES**: ■ Factor **OE**: ◆

Negative pole of Factor **Ex**: ○ Factor **A**: ▽ Factor **C**: △ Factor **ES**: □ Factor **OE**: ◇

The 21 items used as a subdeck in Sample 7 are marked with *.

The smaller pools of data involved in estimating the other dissimilarity matrices reduce their reliability. Nevertheless, the trends were the same: the "positive" / "negative" opposition was universally paramount. In the dendrogram representing the semantic-gradient dissimilarities D^R, there was more intermixing between the A– and ES– clusters, and between the A+ and ES+ clusters. The status of Openness was poorly defined: the OE+ cluster was divided, with one half linked to C+ (as with D^S) while the other items were closer to the Ex+ cluster. The OE– cluster was intact, but closer to Ex– than to C–.

Canonical Correlations

Those canonical-correlation coefficients R_p (p ranging up to 5), which are significant at the $p < .05$ level, are listed in Table 1, for all pairwise comparisons of X^S, X^O, X^R, $X^G{}_5$. Table 1 also shows the R_p from comparing the individual MDS solutions and the 5-dimensional group solution $X^G{}_5$ (combining those three data sets) against X^Q. In general, most pairs agree on three or four mutually-comparable dimensions. $X^G{}_5$ and X^Q are also compared with the semantic-gradient scales y_f and published factor loadings z_f. There is consensus about a fourth dimension, judging from the significant values of R_4 when $X^G{}_5$ is matched against X^S, X^R, X^Q, y_f and z_f. In summary, four dimensions of the five-dimensional MDS group solution are discernible in semantic-gradient scales and in some of the individual solutions. They are readily interpretable (see below) while the fifth dimension is not. We conclude that four dimensions are sufficient to explain most aspects of the observed dissimilarities among words, and the following analyses focus on $X^G{}_4$. Note that the value of R_p does not indicate the *magnitude* of the shared variance within the p-th dimension. The dominant features of $X^G{}_4$ are the first three dimensions, which contribute 39%, 29% and 21% of its total variance, while the fourth dimension is a minor feature, accounting for 11%.

To confirm this agreement of semantic and descriptive information, the peer descriptions from data-collection Sample 6 (using two 27-item subsets) and Sample 7 (using a 21-item subset) were converted into correlation matrices. The results of analyzing these with MDS are lists of coordinates that can be written as two 27-by-4 matrices ($X^{Q1}{}_4$ and $X^{Q2}{}_4$) and a 21-by-4 matrix ($X^{Q3}{}_4$). According to CANCORR analyses, each of these matrices has three mutually-recognizable dimensions in common with $X^G{}_4$. The emergence of a fourth dimension was prevented by the small numbers of items, which impose higher significance thresholds, and also the scarcity of informants: 17, 19, and 24, respectively.

A final MDS solution ($X^T{}_3$) was obtained for 21 items from the triadic data from Sample 8. A maximum-likelihood algorithm was used that directly optimizes the fit to the odd-one-out judgments, avoiding an intermediate step of estimated dissimilarities (Bijmold & Wedel, 1999; Takane, 1978). Given the paucity of the data, only three dimensions were examined. All had their counterparts in $X^G{}_4$ according to CANCORR, though the third did not reach the usual thresholds of significance ($p = .065$).

Table 1
Successive Canonical Correlations R_p ($1 \leq P \leq 5$) Extracted from Pairs of Variable Sets for the 60 Big-Five Markers

(A) Variables are Five-Dimensional Coordinate Matrices. Comparisons among Separate Internal-Data Solutions X^S, X^R, X^O; Separate Solutions against Combined Solution $X^G{}_5$; Internal-Data Solutions against External-Data Solution X^Q

	R_c	R_2	R_3	R_4	R_5
X^S / X^O	.978*	.899*	.633*	.137	.068
X^S / X^R	.951*	.879*	.846*	.317	.101
X^O / X^R	.864*	.819*	.688*	.156	.041
$X^G{}_5$ / X^S	.998*	.983*	.931*	.519*	.183
$X^G{}_5$ / X^O	.978*	.900*	.787*	.316	.050
$X^G{}_5$ / X^R	.998*	.983*	.926*	.834	.738*

* $p < .01$.

	R_c	R_2	R_3	R_4	R_5
$X^G{}_5$ / X^Q	.900*	.778*	.695*	.532*	.275
X^S / X^Q	.863*	.764*	.394	.293	.138
X^O / X^Q	.844*	.546*	.392	.274	.172
X^R / X^Q	.874*	.775*	.665*	.511*	.203

(B) Comparisons of Big-Five Factor-Loading Matrices y, z against $X^G{}_5$ and X^Q

	R_c	R_2	R_3	R_4	R_5
z / $X^G{}_5$	.971*	.964*	.933*	.843*	.706*
z / X^Q	.928*	.830*	.791*	.469*	.200
y / $X^G{}_5$	.982*	.980*	.966*	.889*	.691*
y / X^Q	.920*	.774*	.637*	.530*	.186

* $p < .01$.

MDS

A brief commentary on the features of this combined spatial model complements the description of the dendrogram. We defer discussion of the minor fourth dimension. Between them, the first three dimensions differentiate the clusters of factor markers.

This is illustrated in Figure 2. A salient feature of the MDS solution is that the points were roughly equal in distance from the origin; they are confined to a spherical shell. We exploit this quality, and simplify the problem of displaying the solution, by showing how the points would look from the origin (the center of the sphere). In order to project them onto two dimensions, this perspective ignores their varying radial distances from the center and treats them

as points on the surface of a globe. We divide the globe into two hemispheres (negative- and positive-*D1*) halves, and flatten them separately in Figure 2. The two outer circles indicate the “equator” where *D1* = 0. The center of each circle is one pole of the *D1* axis, while angles of 30º and 60º away from that axis are shown as concentric inner circles.

The sphericity may partly be artefactual, as noisy data have the potential to create internal inconsistencies among the dissimilarities, which in turn lead in two or more dimensions to a kind of annular solution. But the arrangement is plausible, if radial distance from the origin is interpreted as “qualification value” or item specificity (the converse of what Maraun, 1997, termed “integrative centrality”). The items are all drawn from the same lexicon and should all be approximately equally narrow in their application. The origin itself is a point without qualities, a complete vagueness.

The *D1*– hemisphere shows a roughly circular arrangement of items, forming clusters in the sequence (clockwise from top) A– and ES– together, C–, OE–, then Ex–. The clusters of markers at the factors’ positive poles form a symmetrical sequence in the *D1*+ hemisphere, so that the two clusters at the poles of each factor are diametrical opposites.

The weakest distinctions are between the A+ and ES+ clusters, and between A– and ES–. This brings to mind the observation (Peabody & Goldberg, 1989) that the ES factor is relatively labile, in that changes in procedure can shift many of its high-loading scales to measure A instead. We note also the spread of the putative ES markers, especially in ES+, so that *unemotional* and *unexcitable* have migrated across the *D1* midpoint to form a bridge to the Ex– cluster (to be met by *quiet* migrating in the opposite direction)—a bridge reflecting the similar description of the ES+ and Ex– traits in terms of *D2* and *D3*. *Emotional* is also displaced in the opposite direction, from the ES– group in the direction of Ex+.

D4 strengthens the differentiation of the A+ cluster from ES+, and A– from ES–, in a way that Figure 2 cannot show. At the same time, it clarifies the distinctions between C+ and OE+, and C– and OE–. At one extreme of *D4* are traits such as *emotional*, *sympathetic* (A+), *talkative*, *bright* (OE+) and *impractical* (C–), with *unemotional*, *simple* (OE–), *unkind* (A–) and *practical* (C+) at the other extreme. Possible labels for this dimension include “Emotional” or “Abstract / Concrete.”

The alignment of X^G_4 is optimal, in that it accounts for as much as possible of the variation between dissimilarity matrices in terms of relative elongation or compression of the solution along the dimensions. This is because individual-differences MDS postulates that the four matrices were measured in ways that differ from one another by placing more or less weight (salience, importance) on each of the underlying dimensions.

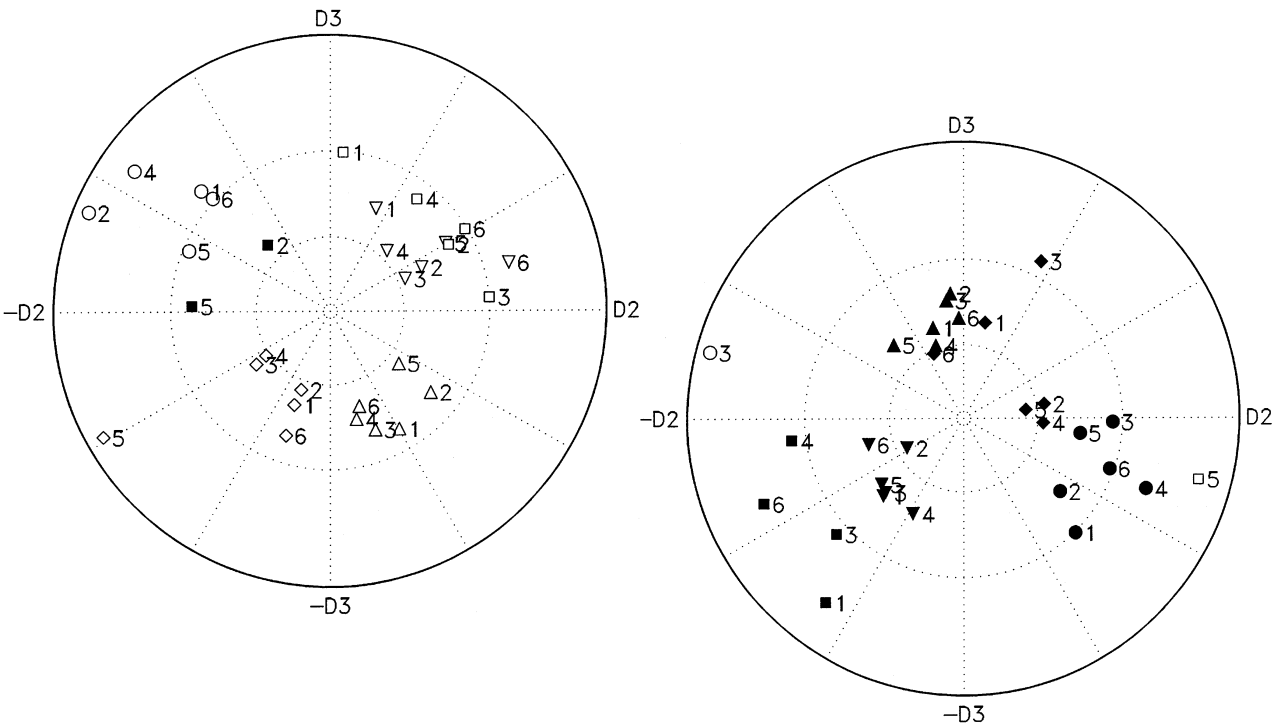

Figure 2. Locations of 60 trait adjectives as determined with MDS, mapped into two dimensions by stereographic projection from a spherical surface. D1– and D2+ hemispheres at left and right, respectively.

Positive pole of Factor **Ex**: ● Factor **A**: ▼ Factor **C**: ▲ Factor **ES**: ■ Factor **OE**: ◆

Negative pole of Factor **Ex**: ○ Factor **A**: ▽ Factor **C**: △ Factor **ES**: □ Factor **OE**: ◇

D1 can plausibly be identified as "Evaluation." In this light, the division of each dendrogram into half-trees appears as a fundamental split between positive and negative Evaluation. The majority of words in this lexicon contain some connotation of value; the markers of Extraversion, plus some words descriptive of Emotional Stability, come closest to being value-free. *D2* distinguishes *bold, energetic, extraverted, demanding* at one extreme from *calm, unexcitable, quiet, introverted* at the other. A possible identification is "Activity." Clearly, there is a strong association with Ex, but as well as the markers of Ex+, "active" traits include those of OE+, ES– and C–, while "inactive" traits include markers of ES+, OE– and C+, as well as those of Ex–. *D3* distinguishes *thorough, systematic, intellectual* from their opposites. It cleanly separates the markers of C+ and OE+ from other positive-Evaluation traits, and those of C– and OE– from other negative-Evaluation traits; possible labels are "Potency" or "Capability," though neither is fully satisfactory.

An item's *D2* and *D3* coordinates together convey the purely descriptive, non-evaluative component of its meaning. Figure 2 can be thought of as a pair of polar coordinate plots, where the angular coordinate represents the ratio between Activity and Potency, while the radial coordinate represents the ratio between these Descriptive components combined, and Evaluation.

The *f*-th semantic gradient y_f can be represented in X^G_4 by a vector, representing the direction in which the values y_{fi} of individual items increase most steeply. The vectors are obtained by stepwise multivariate regression, with y_{fi} as the dependent variable and the items' coordinates x_{ip} as independent variables ($1 \leq p \leq 4$). The multivariate correlations—indicating how well the vectors reproduce the y_f —are R = .96, .97, .95, .91, and .94. On fitting vectors for the published loadings z_f instead, the corresponding correlations are R = .92, .88,0.91, .64, and .84. Note that Ex and ES are the factors represented most and least accurately in the spatial model. This trend remains true when they are represented in the same way within solutions for individual data sets (though the correlations are lower overall).

The components of these vectors on the four dimensions are listed in Table 2. They show Extraversion to be predominantly Active, mildly Good, and slightly non-Potent. Its relatively low *D1* component explains the Ex markers which find their way into the "wrong" half-tree in Figure 1.

Agreeableness is Good but slightly Inactive and non-Potent.

Conscientiousness is Good and Potent (i.e., intellectual evaluation), and Concrete (according to its negative *D4* component).

Emotional Stability is Good, but even more Inactive and non-Potent than Agreeableness.

Openness is Potent and Good, with Abstractness (the positive *D4* component) distinguishing it from Conscientiousness.

Multivariate regression can also be performed in the other direction, predicting x_{ip} (for a given p) from linear combinations of the factor loadings. Using y_f ($Ex \leq f \leq OE$) as the independent variables reproduces the first three dimensions of X^G_4 very well, while *D4* is less amenable: the multivariate correlations are R = .98, .98, .97, and .84. Using z_f as independent variables produces the same pattern, with R = .97, .96, .93, and .81.

Table 2
Item Loadings on the Big-Five Factors as Combinations of their Coordinates in X^G_5, Obtained with Stepwise Multivariate Regression

	R	*D1*	*D2*	*D3*	*D4*
y_{Ex}	.96	.60	.72	–.44	–.14
z_{Ex}	.92	.42	.77	–.41	–.24
y_A	.97	.82	–.38	–.33	—
z_A	.88	.62	–.34	–.42	.21
y_C	.95	.71	–.26	.38	–.33
z_C	.91	.53	–.27	.52	–.36
y_{ES}	.91	.63	–.47	–.40	—
z_{ES}	.64	.22*	–.47	–.25*	-.28
y_{OE}	.94	.80	.26	.37	.20
z_{OE}	.84	.51	.29	.33	.50

* $p < .05$ level. Otherwise, $p < .01$. These are standardized b coefficients.

Discussion

There is a widely-shared framework of expectations about the co-occurrence of personality traits. This is manifested when informants estimate the likelihood with which the presence of one trait predicts another, or furnish comprehensive trait descriptions by extrapolating from inadequate information, or simply rate the similarity of trait words. Significantly, the framework has much in common with empirically-derived models of personality description; in particular, with the Big-Five factor structure and the Five-Factor Model. These could be characterized as a way of formalizing "folk personality psychology."

The impressive degree of convergence makes it less than certain that the empirically-derived models reflect actual regularities in personality variation. An analogy can be drawn with the parallel situation in color vision. FA of color-similarity data deduces five factors while MDS reconstructs a circumplex, the familiar color circle (Guttman, 1966); but this is compatible with a structure of distinct color categories (debate continues as to how far cross-cultural universality extends for the categories, associated with "Basic Color terms"). The same circumplex emerges from *semantic* judgments of similarities among color *names* (Shepard & Cooper, 1992). It is even present in the judgment of color-blind observers, so first-hand experience of all the hue distinctions instantiated by the names is not essential. The point is that a categorical structure, however robust and

well-replicated, is no indication that "redness," "blueness," etc., are properties underlying spectral distributions in the outside world. The regularities giving rise to categories and circumplex alike are those of human visual processing.

By analogy, models derived from trait attribution and semantic judgments relate to the mechanisms and limitations of trait *perception*. The universal nature of the trait structure extracted by FA, with its many cross-cultural regularities, would then reflect shared tendencies to infer and extrapolate; shared limitations in what is observable. There is no need to postulate that a great deal of veridical information about the described personalities is captured in the process. This differs from the "systematic distortion hypothesis" (Shweder & D'Andrade, 1980), because we are not suggesting that observers' preconceptions about trait co-occurrence are actually distorting their observations of behavior.

A more common interpretation reverses the direction of cause and effect, taking the judged similarity between two items to be an accurate report of their co-occurrence in practice, averaged across observed instances of both. This account postulates a neutral memory, untainted by projection, indexed for immediate retrieval of past instances. However, the distorting nature of retrospective memory is well-known (Shweder & D'Andrade, 1980). Dabady, Bell, and Kihlstrom (1999) found no evidence of a salient role for the FFM in organizing personality-trait memory.

Borkenau (1992) proposed that a question about the similarity of two traits is answered by performing a componential analysis of the semantics of both terms, and assessing the extent of overlap. These semantic features also influence the terms' co-occurrence by restricting their simultaneous use. We submit, however, that it is common to use trait labels without any awareness of the features or rules governing their use (unless prompted by anthropologists), and that those rules do not deserve to be treated as fundamental: They are derived from similarities rather than generating them. The course of language acquisition consists of learning new words at such a rate that a child's knowledge of a given word is seldom based on encountering an external-world referent, or memorizing a dictionary definition. Rather, they gain a rough idea of its location in the semantic "map," as hearing it used in different contexts provides its relationship to other words—its near-synonymity to some, its antonymity to others. There is no reason to believe that the trait lexicon is atypical. Indeed, there are no concrete external-world referents for traits. Although they can be "grounded" in terms of observable actions, this is seldom done explicitly.

As was noted above, basing a trait-attribute profile upon inadequate information clearly involves the internal, semantic structure of the trait lexicon, for purposes of extrapolation. We are arguing that even after prolonged observation, the decision whether or not to endorse a trait as descriptive of an individual is just as much a matter or inference and extrapolation, involving the whole pattern of inter-trait similarities (few traits having unambiguous marker behaviors). It may be that a trait profile, however detailed, does not contain a great deal of information, and that is why its gist can be captured by five factors or a three-dimensional vector (for Dabady et al., 1999, the Big Five are "[…] *Blind Date Questions* —the first questions that we would ask about strangers with whom we were about to spend an extended period of time"). Self-descriptions are no different; we are not convinced that there are introspective markers to give an individual privileged access to his or her trait composition.

There is an obvious objection that after millions of years of human evolution, in which a survival advantage accrued to anyone adept at overcoming the observational shortfalls and gaining superior insight into the motives, moods, and personality traits of their fellow-primates, such judgments should actually be quite veridical. But by the same token, concealing one's actual personality traits and dissimulating others have also carried an evolutionary advantage.

Here, it proved possible to represent 60 personality traits as points in a unified spatial model, accounting for their semantic structure, those with similar meanings (according to data from GPA-sorting) being adjacent, and those with antonymous meanings (according to O-sorting) being diametrically opposite. One is reminded of the "...visual representation of [Roget's] *Thesaurus*—a vivid image of words as clusters of star-like points in an immense space" (Osgood, 1971, p. 6). We conclude that MDS is a valid approach to the study of personality descriptors, supplementing the more commonly-used FA.

Our X^G_4 model is four-dimensional, the fourth of less significance. Differences between dissimilarity matrices have been exploited to find the optimal alignment for the MDS solution: the alignment in which such differences are best accounted for in terms of elongation or compression along the axes. The first three axes can be identified with Evaluation, Activity, and Potency, which according to Osgood's (1971) "semantic space" theory, are fundamental qualities or categories, appearing in comparisons of concepts within any semantic domain. These protean qualities can be conveyed by bipolar contrasts such as *good* / *bad* and *nice* / *awful* (for Evaluation), *fast* / *slow* and *noisy* / *quiet* (Activity), and *strong* / *weak* and *deep* / *shallow* (Potency). The fourth dimension is reminiscent of a minor fourth factor (Concreteness or Stability), which Osgood also observed.

For comparison, a market-research study applied sorting procedures and MDS to a lexicon of 88 product-description words (Bimler & Kirkland, 1999). The words were selected to be quite general or metaphorical ones, suitable for describing an imagined consumer of a product or an associated life-style rather than the product directly, and as a result, they had much in common with personality traits. Indeed, that study's lexicon overlaps by 14 words with the one used here. Because that study concentrated on descriptors likely to enhance a product's saleability, a strong Evaluation axis (i.e., a Good / Bad contrast) was not expected. However, of the three dimensions returned from an MDS analysis, *D1* and *D2* were recognizable as Potency and Activity, while

D3—ranging from *stylish, expensive, elegant, chic* to *solid, firm, simple, safe*—corresponds loosely to the *D4* found here.

Mehrabian (1995) presented a similar framework of three factors or scales. He argued that these summarized his personality-description data as adequately, and more parsimoniously, than the scales of the Big-Five framework.

Maraun (1997) used MDS to reduce trait-descriptor correlations to a spatial model, but found *two* dimensions sufficient. However, where we would use two points at diametrical opposites to represent the poles of each bipolar scale, Maraun collapsed them to a single point, and in effect mapped only one hemisphere of the trait domain. In some respects, this simplifies the picture, but it also creates new complications.

Concentrating on the first three dimensions of X^G_4, the observed factors roughly mark out the surface of a double cone. In geometrical terms, the axis of symmetry of the cones is close to *D1*. This replicates a geometrical model put forward to account for trait correlations (Peabody & Goldberg, 1989), though that latter model omits the less-robust ES and OE factors. The axis is tipped slightly, with the effect of bringing A closer to *D1* and Ex closer to *D2*.

We observe that certain pairs of clusters are closer than others; for instance, adjacent clusters in the circumplex sequences shown in the *D1+* and *D1-* hemispheres (Figure 2). Agreeableness is in close proximity to Emotional Stability, and Conscientiousness to Openness, to the extent that one is tempted to conflate them into "social desirability" and "intellectual desirability" factors. Extraversion is the factor coming closest to independence from the others. Continuity in the spatial model suggests that interstitial traits—ones with intermediate ratings or loadings on more than one factor—would be easier to find between factor poles such as Ex+ and OE+, which form neighboring clusters, and harder between poles such as Ex+ and C+, which form more distant clusters. The same applies to clusters separated only by *D1*, as Ex- is separated from ES+, in contrast to a pair such as C+ and OE-.

D1 is an inextricable component of all the factors, in various combinations with the other dimensions. Thus, the factor to which a word is assigned is determined by its evaluative quality as well as its descriptive qualities. *Energetic* and *temperamental*, with comparable locations on the two circumplexes of Figure 2 but opposing *D1* values, are markers of Ex+ and ES- respectively. There are many other such "mirror-image" pairs (useful for contrasting a feature of one's own personality to the same feature in some other person); better examples include the pair *generous* and *spendthrift* (assigned to A+ and C- in Peabody & Goldberg, 1989), or their antonyms *stingy* and *thrifty* (assigned to A- and C+).

Conclusions

A key outcome is the high comparability between internal and external structure, epitomized by X^G_5 and X^Q. The former (derived from three different forms of semantic judgment) is compatible with published factor loadings z_f, which sum up the result of conducting FA upon a large body of peer descriptions (Goldberg, 1992). Moreover, those factor loadings correlate well with "semantic gradients" y_f provided by informants ranking the items along the Big-Five factors purely on the basis of their meanings. Descriptive and semantic judgments would seem to draw upon the same implicit knowledge about the 60 words.

Note that the ability to recover five-factor loadings, prompted by the definitions of factor poles, is a statement about the internal coherence of the semantic structure. It does not imply any special status for the factors themselves: Conceivably, the items could be ranked with equal consistency along non-canonical factors, "falling in the cracks" between the Big Five, such as the axes of Dominance and Nurturance (Trapnell & Wiggins, 1990).

Perceptions of similarities within a lexicon of 60 trait-descriptive adjectives reveal a relatively simple semantic structure. A spatial model was capable of accommodating the range of meaning among the words: or rather, the ranges of meaning, as the space had three major dimensions. We can speculate that this low dimensionality does not indicate any innate simplicity of the lexicon, but instead reflects the cognitive limitations of our informants. It may be difficult to attend simultaneously to more than three forms of difference between items as one considers contributions to their overall dissimilarity.

The use of MDS allows an individual-differences approach, resulting in a spatial configuration that is rotated to a non-arbitrary set of axes (those on which the various data-collection procedures differ). The configuration is dominated by an Evaluation axis, on which most items load to some extent, modulated by three subsidiary axes to produce the classic five factors. Two of these subsidiary axes were identifiable as Activity and Potency. This is a return to quite an old-fashioned model of trait description. As well as Osgood's (1971) work, a similar geometric model was put forward for the single-word trait lexicon by Peabody and Goldberg (1989). *D2* and *D1* are similar to the two robust superfactors obtained at the highest level of hierarchical FA: "[...] something akin to the construct of overcontrol (as compared to undercontrol)", and "[...] something much like the dynamic construct of ego-resiliency (as compared to ego-brittleness), of being adaptively tuned to the surrounding world" (Block, 2001, p. 103). Using slightly different FA methodology, Carroll (2002) obtained a pair of third-order superfactors. The first, "General Social Competence," could be construed as a combination of *D1* and *D2*, while the second, "General Goodness of Personality," would be a combination of *D1* and negative *D2* in the current terms.

Support for the non-arbitrary, non-rotatable status of at least the Evaluation and Potency axes comes from Vonk (1995), where fictitious characters were located within a MDS model as "ideal points," on the basis of trait attributions extrapolated from brief behavioral vignettes. Augmenting the vignettes with Potency-inconsistent information displaced the corresponding ideal points along the Potency axis; adding Evaluation-inconsistent information had no such effect.

Significantly, the external structure displayed in peer-description correlations is seen to be similar to internal structure. Divergence between them is partly attributable to differences in the salience of particular dimensions. Does this mean that the apparent complexity of human personality can actually be reduced to three broad, molar dimensions? We prefer to ascribe the simplicity to the informants' limitations, in this case not in cognition, but in observation.

References

Bijmold, T.H.A., & Wedel, M. (1999). A comparison of multidimensional methods for perceptual mapping. *Journal of Marketing Research, 36*, 277-285.

Bimler, D.L., & Kirkland, J. (1999). Capturing images in a net: Perceptual modelling of product descriptors using sorting data. *Marketing Bulletin, 10*, 11-23.

Block, J. (1961). *The Q-sort method in personality assessment and psychological research*. Springfield, IL: Charles C. Thomas (reprinted 1978, Palo Alto, CA: Consulting Psychologists Press).

Block, J. (2001). Millennial contrarianism: The five-factor approach to personality description 5 years later. *Journal of Research in Personality, 35*, 98-107.

Borkenau, P. (1992). Implicit personality and the five-factor model. *Journal of Personality, 62*, 295-327.

Carroll, J.B. (2002). The Five-Factor personality model: How complete and satisfactory is it? In H. Braun, D.N. Jackson, & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 97-126). Mahwah, NJ: Erlbaum.

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

Dabady, M., Bell, M., & Kihlstrom, J.F. (1999). Person memory: Organization of behaviors by traits. *Journal of Research in Personality, 33*, 369-377.

D'Andrade, R.G. (1965). Trait psychology and componential analysis. *American Anthropologist, 67*, 215-228.

Davison, M.L., & Skay, C.L. (1991). Multidimensional scaling and factor models of test and item responses. *Psychological Bulletin, 110*, 551-556.

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

Graziano, W G., Jensen-Campbell, L.A., Steele, R.G., & Hair, E.C. (1998). Unknown words in self-reported personality: Lethargic and Provincial in Texas. *Personality & Social Psychology Bulletin, 24*, 893-905.

Guttman, L. (1966). Order analysis of correlation matrices. In R.B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 438-458). Chicago: Rand McNally.

Hakel, M.D. (1969). Significance of implicit personality theories for personality research and theory. *Proceedings of the 77th Convention of the APA, 4*, 403-404.

Harris, R.J. (1975). *A primer of multivariate statistics*. New York: Academic Press.

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

Johnson, J.A., & Ostendorf, F. (1993). Clarification of the Five-Factor Model with the Abridged Big Five Dimensional Circumplex. *Journal of Personality & Social Psychology, 65*, 563-576.

Kruskal, J.B., & Wish, M. (1978). *Multidimensional scaling*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-011. Beverly Hills, CA: Sage.

Maraun, M. D. (1997). Appearance and reality: Is the Big Five the structure of trait descriptors? *Personality & Individual Differences, 22*, 629-647.

Mehrabian, A. (1995). Relationships among three general approaches to personality description. *Journal of Personality, 129*, 565-581.

Osgood, C.E. (1971). Explorations in semantic space: A personal diary. *Journal of Social Issues, 27*, 5-64.

Passini, F.T., & Norman, W.T. (1966). A universal conception of personality structure? *Journal of Personality & Social Psychology, 4*, 44-49.

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

Shepard, R., & Cooper, L. (1992). Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science, 3*, 97-103.

Sherman, R.C. (1972). Individual differences in perceived trait relationships as a function of dimensional salience. *Multivariate Behavioral Research, 7*, 109-129.

Shweder, R.A., & D'Andrade, R.G. (1980). The systematic distortion hypothesis. *New directions for methodology of social and behavioral science, 4*, 3-38. San Francisco: Jossey-Bass.

Sneed, C.D., McCrae, R.R., & Funder, D.C. (1998). Lay conceptions of the five-factor model and its indicators. *Personality & Social Psychology Bulletin, 24*, 115-126.

Takane, Y. (1978). A Maximum Likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent – Theory. *Japanese Psychological Research, 20*, 7-17.

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

Vonk, R. (1995). Effects of inconsistent behaviors on personal impressions: A multidimensional study. *Personality & Social Psychology Bulletin, 21*, 674-685.

Wiggins, J.S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

Received May 9, 2006
Revision received October 22, 2006
Accepted November 30, 2006

Appendix A

Let c_{kij} stand for the proportion of sorting partitions, in the *k*-th sorting sequence, in which the *i*-th and *j*-th items are placed in separate groups. The initial arrangement of items into piles counts as one partition; for participants who performed GPA-sorting, the finer subdivision into subgroups is another; each of the subsequent arrangements from the Addition stage contributes another partition. Then the average value of c_{kij} provides an estimate of the dissimilarity between items *i* and *j*:

$$d^{S}_{ij} = \sum_k c_{kij} / K_{ij}$$

Here, K_{ij} is the number of sorting sequences in which those items are both involved, and the index *k* ranges only over those cases. Note that K_{ij} is *not* the number of subjects, because many subjects provide more than one sequence, while many sequences lack one or other items, or both. A large K_{ij} means that d^{S}_{ij} is known within tighter confidence bounds, being less subject to sampling error. This could be incorporated in the MDS analysis by weighting the dissimilarity estimates by their reliability—an option available in some MDS packages—but this refinement was not pursued here.

Appendix B

An estimate of the dissimilarity between the *i*-th and *j*-th items is the number of times they are in piles chosen as "opposite," as a proportion of the opportunities they have for this to happen. Let q_{ij} stand for the number of opposite-pile selections involving items *i* and *j*, summed over participants, and o_{ij} for the number of such selections in which those items belong to the opposite piles. Items are only "involved" if: (a) they are both present in that participant's deck of cards, and (b) they belong to different piles, and (c) they are not in piles which have already been picked out as opposites and put to one side. Then:

$$d^{O}_{ij} = o_{ij} / q_{ij}$$

If q_{ij} is zero, d^{O}_{ij} is treated as "missing data."