European Journal of Personality, Vol. 11:1, 15–31 (1997)

A Comparison of Big-Five Structures of Personality Traits in Dutch, English, and German

WILLEM K. B. HOFSTEE*, HENK A. L. KIERS and BOELE DE RAAD
University of Groningen, The Netherlands

LEWIS R. GOLDBERG
University of Oregon and Oregon Research Institute, USA

and

FRITZ OSTENDORF
University of Bielefeld, Germany

Abstract

We compare Big-Five factor structures found in Dutch, American English, and German, and present a joint structure. The data consist of self- and peer ratings of 600 subjects with 551 Dutch trait-descriptive adjectives, 636 subjects with 540 English adjectives, and 802 subjects with 430 German adjectives. On the basis of 126 common items, we assess the congruences between the factors as originally published, as resulting from target rotations, and from simultaneous rotations. With the exception of the Dutch Factor V, the Big-Five factors recur across languages in a relative but not in a strict sense. Moreover, at a more detailed level differences in the positions of the axes are uncovered. By applying a split-sample technique to the three data sets, we verify that these differences do not arise through unreliability. Also, few trait terms appear to have the same precise meaning across these three languages; such labels therefore cannot serve as anchor concepts for an international language of personality. © *1997 by John Wiley & Sons, Ltd.*

INTRODUCTION

Can we discuss personality differences across language borders? For example, can the first author, who is Dutch, meaningfully say in English to the fifth author, who is German, that the fourth author, who is American, is the most extraverted of the five authors? At another level, what is meant by the hypothesis that the Big-Five (see e.g. Digman, 1990; John, 1990) factors of personality recur across languages?

*Correspondence concerning this article should be addressed to Willem K. B. Hofstee, Department of Psychology, University of Groningen, Grote Kruisstraat 2-I, 9712 TS Groningen, The Netherlands.

CCC 0890–2070/97/010015–17$17.50

Received 28 February 1996
Accepted 30 August 1996

In an objective (as opposed to the present judgmental) approach to individual differences it would seem that one *can* say, across language barriers, that Jan is more intelligent than John. However, if the concept of intelligence differs systematically—as it probably does to some extent even across Western languages or cultures (see Neisser *et al.*, 1996)—then the comparison becomes tenuous even at the objective level. *A fortiori*, judgments of personality are problematic.

What is wrong with local definitions of personality concepts? One would have to accept the ultimate consequence of admitting idiosyncrasies. Hardly any two individuals agree precisely with respect to any concept; the relative lack of agreement on trait terms is evident from studies comparing the internal structures of traits held by different respondents (Peabody, 1984). Nonetheless, trait concepts are needed in clinical diagnosis, in personnel selection, in job counselling, and in many other everyday situations. One task of science is to sharpen these concepts and to promote their common understanding, not only between individuals but also across language borders; and to do so without going to the extreme of arbitrary operational definitions that would isolate the scientific community from the applied professions and everyday discourse. Even if one wishes to focus on the differences between concepts, that cannot be done without a common base.

Can dictionaries serve to pin down the meaning of trait terms? Goldberg and Kilkowski (1985) have shown that to some extent they can. Mainly, however, lexical sources provide documentation of the shiftiness of meanings. For example, here is one of the many synonym paths from Good to Bad that can be derived from Rodale's (1961) thesaurus: *Good*–Considerate–Circumspect–Wary–Suspicious–Envious–Malicious–*Bad*.

Essentially the same effect may be demonstrated between two languages by listing the translations of a trait term, their back-translations, the translations of these back-translations, and so forth. Words have many uses. Direct synonyms have partially overlapping sets of uses. (The limiting case of strict synonymity hardly ever occurs.) Indirectly, just about all words of a certain class, like trait adjectives, are connected by synonymity.

To solve the problem of intrinsic shiftiness, linguists typically encourage contextual specification. However, the search for basic dimensions of personality is not served by the inflation of the specific variance that would result from such conditionings. The appropriate approach is psychometric, and it consists of uncovering common meaning and discarding specific variance.

If single trait terms provide insufficient leverage for cross-national understanding, can factors or clusters do the job? One might question how factors or clusters can be said to correspond, without assuming that any trait term has an exact translation. The answer is provided by the psychometric principle of aggregation. Factors are weighted averages of variables, and are thus more robust than single terms. With more than one factor, a simultaneous rotation procedure will ensure an optimal fit. In this manner, factors can be matched even though no single term or limited set of terms is accepted as exactly equivalent.

Like other studies comparing the Big Five across languages (see e.g. Bond, 1979; Ostendorf and Angleitner, 1994; for an overview, see Katigbak, Church and Akamine, 1996), we use the psychometric approach. However, rather than starting from a particular set of traits or questionnaire items in one language and translating them into another (the 'imposed-etic' strategy; see Berry, 1969; Katigbak *et al.*,

1996), we take an 'emic' approach and compare independent Big-Five structures. Investigating whether a particular questionnaire or trait list works when translated into another language cannot answer the question of whether the measured constructs would have come up in the other culture by themselves. In order to investigate the cross-cultural replicability of the Big-Five factors of personality, as we set out to do, it is insufficient to demonstrate that factors match upon translation: there is no guarantee that a factor is prominent enough to be among the Big Five in the other culture.

The complications brought about by the emic approach are threefold. First, there may or may not exist systematic differences among language communities in the distribution of personality traits. Second, language *corpora* may differ: certain trait dimensions may have more terms in one language than another and may therefore explain more variance in factor-analytic solutions. Third, the selections made by the national research teams may (and do) differ. In selecting trait terms from the lexicon, decisions have to be made at the strategic (e.g. whether or not to include specific abilities such as Musical, or strongly evaluative traits such as Good and Bad) and at the executive level (e.g. deciding whether Inspiring is a trait or merely a social effect). Both types of decision are matters of judgment, and are thus subject to variation. In the present study, the selection of terms is all the more important because of the additional selection criterion of representativeness in forming subsets, which may have meant different things to the research teams. There is no sure way of isolating personality and *corpus* differences from differences arising through item selection.

Apart from this major complication, some minor ones should be considered. First, differences in data-analytic procedures may detract from the cross-language correspondence between factor solutions. We meet this problem by comparing both the Big-Five structures as originally published and the structures that arise when the data are processed in the same manner. Second, lack of correspondence can be attributed in part to unreliability of the single solutions. We introduce a method to estimate these unreliabilities, so that cross-language congruence coefficients may be compared to intra-language congruences.

As a framework for factor interpretation, we use the Abridged Big-Five Dimensional Circumplex (AB5C) trait taxonomy (Hofstee, De Raad and Goldberg, 1992). This model represents traits by their primary and secondary loadings, for example, Responsible as III + II +. Likewise (as a factor is a variable among other variables), if the Dutch Factor III has congruence coefficients of 0.8 and 0.3 with the American-English Factors III and II, respectively, that Dutch factor is III + II + in the American-English framework. The AB5C model thus allows for more nuanced comparisons than the simple-structure approach, where factors are equated by their highest congruence coefficients. The pairings of the ten poles of the Big-Five dimensions create 90 facets, including the ten factor-pure clusters (but not the ten combinations of the positive and negative poles of the same factor, e.g., I + I −, which are impossible).

The use of the AB5C model further enables us to test a specific aspect of the lexical hypothesis, which holds that salient individual differences give rise to trait terms. There are more traits associated with some blends of factors (e.g., I + II +) than with other blends (e.g., I + II −); both the American-English (Hofstee *et al.*, 1992) and Dutch (Hofstee and De Raad, 1991) AB5C structures of trait adjectives show an uneven spread of traits across facets. Assuming that the factor positions can

be made to correspond, and assuming similar trait distributions in the populations, a corollary of the lexical hypothesis is that the distribution of traits over facets is similar across languages. We test this corollary with the present data.

METHOD

Adjectives

Three sets of trait-descriptive adjectives were used: 551 Dutch adjectives described by Hofstee and De Raad (1991), 540 English adjectives described by Hofstee et al. (1992), and 430 German adjectives described by Ostendorf (1990). In the Dutch set, there are relatively few terms pertaining to intelligence and capacities, as a consequence of a selection criterion that emphasized temperamental traits; in the German set, intelligence and capacity terms were expressly included. Apart from these differences, each set is supposed to be representative of the personality-descriptive adjectives in that language.

Translations

The first and third authors undertook the cross-translation of the Dutch and English terms. Using standard Dutch–English and English–Dutch dictionaries, they noted all English translations of the Dutch set of adjectives that were found in the English set, and all Dutch translations of the English adjectives that were found in the Dutch set. Subsequently, the best one-to-one translations were selected. As this selection process is inevitably judgmental to some extent, we present an extensive illustration.

The alphabetically first term in the Dutch set was Aalglad, the only translation of which in the English set was Glib. The only translation of Glib in the Dutch set was Aalglad, so in this case the selection was straightforward. In total, there were 97 such cases. About half of these consisted of fairly difficult and behaviourally specific cognate pairs, like Asocial, Contemplative, Diplomatic, Exhibitionistic, Individualistic, and Masochistic.

The second Dutch term was Aanhankelijk, with a single translation of Affectionate. Affectionate, however, had two translations, Aanhankelijk and Teder. The sole back-translation of Teder was Affectionate. We judged Teder to be the better of the two translations of Affectionate. Naturally, the reasons for such judgments varied from case to case. In this instance, we reasoned that Teder is the more adult version of Affectionate, and is thus more likely to be applicable to subjects from the present samples. Thus, Aanhankelijk was lost in the one-to-one translation process. In general, our considerations were psychological as well as semantic. In no case, however, did we inspect the patterns of factor loadings in the two languages.

For the third Dutch term, Aanmatigend, the most obvious translation in English is Arrogant; however, Arrogant was not in the English set. Indeed, if it had been included, it would have been assigned to its Dutch cognate (which was included in the Dutch set). The only available translation for Aanmatigend was Pretentious, which had been paired with its Dutch cognate. So Aanmatigend also was lost. More generally, in the interest of objectivity automatic priority was given to cognate

translations, even though such translations are not necessarily the best ones in all psychological respects.

The fourth Dutch term was Aanstellerig, translated as Affected, which was not in the English set. No other available translations were listed in the dictionary. However, in the process of our intensive back-and-forth searches we hit upon another single term, Theatrical, in the English set, whose cognate translation was not in the Dutch set. We decided to accept the two as each others' translations. The motive here was maximization of the number of pairs: as there is no sharp boundary between exactly translatable and untranslatable terms, we sought power in larger numbers of pairs.

The fifth Dutch term had a one-to-one translation. The sixth, Aartsgierig, had a single translation, Stingy, which had three back-translations: Gierig, Krenterig, and Vrekkig. Together, these three produced one more English translation, Miserly, which in turn translated as Gierig and Vrekkig. Thus, we had a closed system comprising four Dutch and two English adjectives, out of which two pairs could be formed. We decided to retain Gierig–Stingy and Vrekkig–Miserly, the two most literal translations.

The seventh term in the Dutch set was Accuraat, the cognate of which was not in the English set; its only available translation was Precise, which had been taken by its Dutch cognate. So Accuraat dropped out. Another translation of Precise was Nauwgezet, which illustrates a final strategic problem to be discussed here—that of dealing with endless ramifications. Nauwgezet had four translations in the English set (Conscientious, Particular, Scrupulous, and Strict); together, they produced seven new Dutch back-translations (Conscientieus, Gewetensvol, Plichtsgetrouw, Kieskeurig, Veeleisend, Streng, and Stipt); this set produced ten new English back-translations (Fastidious, Finicky, Demanding, Exacting, Austere, Rigid, Stern, Tough, Punctual, and Prompt), which produced nine new Dutch ones (Pietepeuterig, Nuchter, Sober, Onbuigzaam, Star, Hardvochtig, Hard, Punctueel, and Vlot). Then the numbers went down: six new English terms (Down-to-earth, Restrained, Callous, Hard, Harsh, and Sociable), four new Dutch ones (Ongevoelig, Bikkelhard, Glashard, Keihard), one new English term (Insensitive), and one Dutch term (Gevoelloos), after which no new terms were produced. Leaving out three pairs of cognates (Conscientious, Hard, and Punctual), a set of 20 Dutch terms and 18 English terms remained. Obviously, there is a very large number of ways to select a maximum of 18 pairs. We therefore curtailed such sets, by excluding any terms whose meanings were too far removed from the starting term. The above example is by no means unique or even extreme: without some curtailment, the majority of terms would have been elements in such amorphous and drifting sets. With more comprehensive dictionaries than the standard ones that we used, most terms might have come together in one gigantic set.

To sum up this phase in the translation process, judgment entered into curtailing the sets of terms, and in selecting the pairs from these sets. Judgment in the pairing was restricted by automatically accepting cognate pairs, of which there were 84, and more generally by preferring literal translations. The pairings could have been further optimized by consulting the patterns of factor loadings in the two languages, but that would have led to spurious outcomes. (For each English term, one could select the Dutch term whose factor loadings correspond most highly, and subsequently conclude that the factors correspond almost perfectly.) Judgment in

curtailing the sets may have influenced the outcome for some terms that were at the semantic edge of a set and received their translation from within, but would have been better off in the context of another set. However, we were aware of that possibility: members of already-translated pairs would sometimes pop up as translations of terms in a new set, and these terms were then reconsidered. With different translaters, and with a different order of terms (for example, reverse alphabetical), the result might have been slightly different. In view of the labour intensity of the procedure, however, we did not perform such a reliability check.

The translation process sheds some light on subtle differences in the composition of the Dutch and English sets of terms. In total, 111 of the 551 Dutch terms (20%) had no translation in the English set; for 164 of the 540 English terms (30%) no Dutch translation was available. We illustrate these differences with cognates that are missing in one of the two sets. For example, one could say in Dutch that a person is Dociel, Eloquent, Feminien, or Serviel, but words like these are probably much more infrequently used in Dutch than in English, and are frowned upon as being Latinisms by many. The reverse also occurs, though probably with lower frequency; examples are Laconic, Infantile, Labile, and Recalcitrant, which are missing from the English set.

More interesting, however, are missing cognates that may reflect a systematic difference in the composition of the sets. The Dutch set contains a number of extremely negative terms: Asociaal, Crimineel, Despotisch, Destructief, Dictatoriaal, Hysterisch, Immoreel, Narcistisch, Neurotisch, Pervers, Sadistisch, Schizofreen, and Tyranniek, which were not included in the English set. The English set included a few social-effect terms such as Attractive, Dramatic, Magnetic, and Transparent, and a few attitude terms like Liberal, Progressive, Provincial, Religious, and Traditional, which had no counterparts in the Dutch set. However, none of these discrepancies between the sets diminishes the potential of the common subset to capture Big-Five variance, as the terms outside the intersection are not central in defining the trait space.

The number of Dutch–English pairs that resulted was 275, which is about half the number of items in the Dutch and English sets. Subsequently, the fifth author, in consultation with the first and third, selected the best German term, if present, among the set of 430 for each Dutch–English pair, using a procedure that was similar to the one described above. Satisfactory triplets could be constructed in 126 cases. If it is assumed that a German–Dutch and a German–English translation would have succeeded for about half of the German items (as in the Dutch–English case; see also De Raad, Perugini and Szirmák, 1996), and if it is further assumed that the two selections would have been uncorrelated, the expected size of the triplets set would have been $\frac{1}{2} \times \frac{1}{2} \times 430$, which is about the same as the observed number.

Such diminishing numbers led De Raad et al. (1996) to restrict their study of five languages to pairwise comparisons, as the number of pentads would have become altogether too small. However, with the pairwise approach an integral solution can no longer be reached: the factors for a particular language differ per comparison.

Subjects

Data from earlier studies were used. The Dutch sample of 600 ratings was used by Hofstee and De Raad (1991). It consisted of 200 pairs of Dutch students (139 of mixed sex, plus 27 male and 34 female pairs), half of whom rated themselves, the other half rating the other member of the pair; plus 200 other students (57.5%

female) who rated themselves. The American sample of 636 ratings was used by Hofstee et al. (1992). This data set consisted of 320 college students (63% female) rating themselves, and 316 of these subjects rating a friend of the same sex and age. The German sample of 802 ratings was used by Ostendorf (1990). In this community sample (mean age, 32.6; 58% female), 414 target persons were rated by themselves and by a close acquaintance; 26 ratings were discarded because of missing values.

Thus in all three samples, self- and peer ratings were used. To correct for differences in means and standard deviations, particularly between self- and peer ratings, each row vector (rater–target combination) was transformed into standard scores.

Analyses

The present study first compares published five-factor solutions in the three languages. In the German (Ostendorf, 1990) and Dutch (Hofstee and De Raad, 1991) studies, the Big-Five solutions are varimax rotations of the first five principal components. In the American study, the positions of the five factors were determined by a principal-components analysis and varimax rotation of a subset consisting of Goldberg's (1992) 100 unipolar Big-Five marker variables (see Hofstee *et al.*, 1992, pp. 147–8), which in turn were based on a wide variety of previous English-language analyses. The input for the first part of the present study thus consisted of the 551×5 matrix of factor loadings of Dutch adjectives, the 540×5 matrix of loadings of English adjectives, and the 430×5 matrix of loadings of German adjectives. Of these matrices, the 126 triplets of translations form a common subset.

Additional analyses starting from the raw data sets were carried out to investigate to what extent any lack of cross-cultural correspondence could be attributed to differences in data-analytic procedures and to unreliability of the factor structures.

RESULTS

Congruences for the original factors

Congruence coefficients (Tucker, 1951) calculated between all pairs of factors over the 126 trait variables are provided in Table 1. For Factors I–IV, congruence was highest for the corresponding factors in all three (Dutch–English, Dutch–German, and English–German) comparisons. For the Dutch Factor V (V_d), congruence was higher with the English Factor I_e (0.56) than with Factor V_e (0.41). Thus in American-English terms, V_d is $I_e + V_e +$; marker terms for that facet are Expressive, Adventurous, and Dramatic *versus* Passive, Meek, and Dull (see Hofstee *et al.*, 1992, p. 157). Similarly in the Dutch–German comparison, the congruence between V_d and I_g was 0.61, whereas the congruence between V_d and V_g was only 0.37; this latter value was also lower than the congruence between IV_d and V_g (0.46). With these three exceptions all pertaining to Factor V_d, the analysis provides empirical evidence of the relative correspondence of the Big-Five factors in the three languages.

However, strict correspondence in the sense of congruence coefficients of at least 0.85 (see Haven and Ten Berge, 1977, for this criterion) was not found. Specifically,

Table 1. Congruence coefficients among the Big-Five Factors in the three languages

	I_d	II_d	III_d	IV_d	V_d	I_e	II_e	III_e	IV_e	V_e
I_e	**0.74**	−0.34	−0.16	0.29	**0.56**					
II_e	**0.31**	**0.76**	**0.31**	−0.24	−0.05					
III_e	−0.05	0.23	**0.83**	**0.36**	−0.11					
IV_e	0.18	**0.40**	0.22	**0.62**	0.11					
V_e	0.22	0.02	0.06	0.19	**0.41**					
I_g	**0.80**	−0.18	−0.26	0.21	**0.61**	**0.80**	0.07	−0.19	0.17	0.27
II_g	0.20	**0.84**	**0.35**	−0.17	−0.07	−0.24	**0.82**	0.25	**0.38**	0.12
III_g	−0.02	0.21	**0.73**	**0.43**	0.02	0.02	0.15	**0.75**	**0.44**	0.02
IV_g	0.08	0.04	−0.06	**0.59**	−0.19	0.12	−0.14	0.06	**0.58**	−0.06
V_g	0.24	0.04	0.17	**0.46**	**0.37**	0.23	0.07	0.27	**0.43**	**0.79**

pure matches rarely occur, as will be clear from inspection of Table 1. For example, the German Factor III_g is $III_e + IV_e +$ in American-English terms. The most representative term for that facet is Thorough, which is reminiscent of stereotypical German *Gründlichkeit*. Dutch Extraversion I_d is $I_e + II_e +$ (Sociable) rather than just $I_e +$ (Talkative) in American-English terms. (For these facet markers, see Hofstee *et al.*, 1992, Table 1; there is no reason in principle to use the American-English solution as a reference space here, but it is the most accessible and understandable to all readers).

Target rotations

Orthogonal target rotations, presented in Table 2, using the configuration of each language in turn as a target, were carried out to maximize correspondence. Only three rotation matrices are presented, as the remaining three are transposes thereof. Here, the values on the diagonals are consistently the highest.

Even in Table 2, however, non-negligible ($\geqslant 0.30$) off-diagonal values occur. For example, to obtain an optimal approximation to I_d, considerable weight (0.45) must be given to II_e in addition to I_e, meaning that I_d is more Agreeable than I_e. Conversely, to approximate I_e, a negative weight (-0.38) should be given to II_d. II_d is less Extraverted and more Emotionally Stable than II_e (weights of -0.38 and 0.48 for I_e and IV_e, respectively). III_d does not deviate much. IV_d is less Agreeable (a weight of -0.41 for II_e) than IV_e. Between the Dutch and German solutions, the largest discrepancies are in Factors IV and V, with IV_d more Intellectual and V_d less Emotionally Stable than their respective German counterparts. Note that this discrepancy does not occur between the Dutch and English solutions. No large off-diagonal entries are found in the German–English rotation matrix.

The target rotations confirm that the positions of the factors are somewhat different, even though the matchings are evident. Part of these differences, most notably the rotation of Factors IV and V between the Dutch and German solutions, may follow from systematic differences in item selection. Other differences may reflect the unreliability of factor positions even with sizeable samples such as these. Finally, some of the differences may reflect a differential cultural emphasis.

Table 2. Rotation matrices using the Dutch, English, and German solutions as targets

	Dutch solution					English solution				
	I_d	II_d	III_d	IV_d	V_d	I_e	II_e	III_e	IV_e	V_e
I_e	**0.86**	−0.38	−0.05	0.19	0.29					
II_e	0.45	**0.77**	0.19	−0.41	−0.02					
III_e	−0.01	−0.10	**0.96**	0.25	−0.11					
IV_e	0.03	**0.48**	−0.18	**0.86**	−0.05					
V_e	−0.25	0.14	0.12	0.01	**0.95**					
I_g	**0.90**	−0.20	−0.11	−0.08	−0.38	**0.96**	0.22	−0.15	0.06	0.05
II_g	0.26	**0.93**	0.13	−0.20	−0.05	−0.23	**0.93**	−0.02	0.28	0.05
III_g	0.09	−0.08	**0.93**	**0.35**	−0.05	0.15	−0.00	**0.95**	0.22	−0.18
IV_g	0.15	0.14	−.32	**0.75**	−0.54	−0.02	−0.27	−0.25	**0.90**	−0.24
V_g	−0.32	0.25	−0.10	**0.51**	**0.75**	−0.01	−0.13	0.13	0.25	**0.95**

Simultaneous rotations

To arrive at an optimal matching of the factors without favouring one of the three positions, simultaneous rotations of the three matrices of factor loadings were carried out. For this purpose we used Generalized Procrustes Analysis (see e.g. Ten Berge, 1977), by which the sum of squared differences between corresponding elements of all pairs of rotated loading matrices is minimized. As a second step, the average of the rotated matrices was rotated by means of varimax, and this second rotation was also applied to the individual rotated loading matrices. The latter procedure, proposed by Hakstian (1973), rotated the loading matrices to a simple-structure cross-national position. These rotated factors will be denoted as cross-national factors. Table 3 shows the improvement in fit when the three independent solutions were rotated to a joint solution.

Table 4 provides the rotation matrices for transforming the national factors into cross-national factors. In all cases, the diagonal values are the highest in their row and column, confirming again the relative generality of the Big-Five factors. Nonetheless, the off-diagonal values in Table 4 are of interest in their own right. For the Dutch solution, the most sizeable off-diagonal value occurs in the $I \times V$ plane, turning the original Factor I_d towards the original Factor V_d. Slight rotations take place in the other planes: (i) in the $I \times II$ plane the original Factor I_d becomes somewhat less Agreeable, and Factor II_d becomes more Extraverted; (ii) in the $II \times V$ plane, V_d becomes more Agreeable; (iii) in the $III \times IV$ plane, Emotional Stability is added to III_d and IV_d becomes a little more Disorganized; and (iv) in the $IV \times V$ plane, V_d moves closer to Intellect.

The most conspicuous rotation in the English configuration takes place in the II x IV plane, with Factor II_e (originally Warmth) moving in the direction of Trust, and IV_e (originally defined negatively by Moodiness and Jealousy) moving toward Irritability and Temperamentalness. Minor rotations take place in the $I \times II$ plane, making the original Extraversion factor somewhat more Sociable, and the original Warmth factor somewhat more Agreeable; and in the $IV \times V$ plane, moving IV_e in the Imperturbable direction and bringing V_e somewhat closer to Controlled Intelligence (Peabody and Goldberg, 1989).

Table 3. Congruence coefficients betwen corresponding factors in different languages before and after rotation to a joint solution

	Factor					
	I	II	III	IV	V	Mean
Dutch/English						
Before rotation	0.74	0.76	0.83	0.62	0.41	0.67
After rotation	0.87	0.85	0.86	0.73	0.42	0.75
Before rotation*	0.81	0.82	0.83	0.57	0.37	0.68
After rotation*	0.85	0.83	0.85	0.64	0.44	0.72
Dutch/German						
Before rotation	0.80	0.84	0.74	0.59	0.37	0.67
After rotation	0.87	0.87	0.80	0.77	0.41	0.74
Before rotation*	0.81	0.85	0.77	0.60	0.40	0.68
After rotation*	0.87	0.83	0.80	0.73	0.44	0.73
English/German						
Before rotation	0.80	0.82	0.75	0.58	0.79	0.75
After rotation	0.84	0.85	0.78	0.72	0.83	0.80
Before rotation*	0.83	0.82	0.77	0.52	0.71	0.73
After rotation*	0.85	0.83	0.79	0.72	0.74	0.79

Note: *Upon reanalysis using the raw data.

The rotation matrix for the German solution is almost an identity matrix. Thus the German solution appears to take an intermediate position between the Dutch and English solutions.

The cross-national AB5C solution

For a substantive interpretation of the cross-national factor structure, the Abridged Big-Five Dimensional Circumplex (AB5C; see Hofstee *et al.*, 1992) is an appropriate tool. After rotation to the joint structure, the 126 terms in each of the three languages were classified according to this algorithm. The results at the trait level show that few adjectives have the same precise meaning even across these three closely related languages. Most notably, complete correspondence for both primary and secondary loadings across all three languages occurred in only ten of the 126 triads. Between Dutch and English, 32 adjectives (25%) had corresponding primary and secondary loadings, as compared to 25 adjectives (20%) between Dutch and German, and 33 adjectives (26%) between English and German.

A more liberal criterion for correspondence is provided by the angles between trait vectors in the five-dimensional space. Setting the criterion at 30° or less, 58 terms (46%) correspond between Dutch and English, 62 (49%) between Dutch and German, and 69 (55%) between English and German. Complete correspondence for all three languages occurs in 34 of the 126 triads (27%). Within the subset of mostly literal translations, one may conclude that the probability that a dictionary gives the psychologically appropriate translation of a trait-descriptive adjective is about 0.5

Table 4. Rotation Matrices to a Cross-National (C) Position

	I_c	II_c	III_c	IV_c	V_c
I_d	**0.90**	0.29	−0.01	−0.01	**−0.32**
II_d	−0.21	**0.94**	−0.04	0.07	0.25
III_d	−0.03	0.07	**0.96**	−0.26	−0.05
IV_d	0.15	−0.10	0.27	**0.90**	0.29
V_d	**0.34**	−0.13	−0.03	−0.34	**0.87**
I_e	**0.98**	−0.19	0.02	0.01	−0.05
II_e	0.18	**0.90**	0.04	**−0.38**	−0.06
III_e	−0.03	−0.06	**1.00**	−0.06	0.00
IV_e	0.07	**0.38**	0.07	**0.90**	0.20
V_e	0.04	−0.03	−0.02	−0.21	**0.98**
I_g	**0.99**	0.03	−0.11	−0.03	0.01
II_g	−0.03	**0.99**	0.03	−0.13	0.06
III_g	0.12	−0.01	**0.98**	0.14	−0.10
IV_g	0.01	0.14	−0.15	**0.97**	−0.13
V_g	0.01	−0.04	0.08	0.15	**0.99**

between two languages, and across three about 0.25. Considering only the primary loadings, 76 adjectives (60%) correspond between Dutch and English; 87 (69%) between Dutch and German; and 91 (72%) between English and German. Across all three languages, 69 adjectives (55%) correspond. So, even by this crude criterion, the translatability of trait items is by no means perfect.

Table 5 lists the 21 triads of cognate terms with their English spelling. Projected against the total correspondence, the set of cognates is fairly representative: the common root does not result in systematically better translations. Of the 21 triads of cognates, only Energetic (I + III +) and Humorous (I + II +) may safely be directly translated. Dutch and English share, moreover, Egocentric and Egotistical (both II − III −), Hard (II − IV +), Intolerant (II − IV −), and Orderly (III + III +); Dutch and German share Extraverted (I + III −), Generous (I + II +), Humourless (I − V −), Impulsive (I + IV −), and Punctual (III + III +); and English and German share Dominant (I + II −), Intelligent (V + V +), Introverted (I − I −), and Sarcastic (II − III −). Different for all three languages are Creative, Emotional, Naive, Rational, and Tolerant. Generally, the findings raise a *caveat* against overconfidence in the translatability of such terms.

Extension to the full sets of traits

The factor structures of the 551 Dutch terms, the 540 English terms, and the 430 German terms were rotated to the common position, using the rotation matrices that resulted from the simultaneous rotation of the 126 triads. The AB5C procedure (Hofstee *et al.*, 1992) was applied to the matrices of rotated loadings. Table 6, in which the *columns* represent the *primary* loadings and the *rows*, the *secondary* loadings, lists the terms in each facet that have the highest projections in that facet, with a minimum of 0.4. A complete list is available from the authors.

Table 5. AB5C Locations of Cognate Trait terms

	Dutch	English	German
Creative	5+4+	5+5+	5+1+
Dominant	2−5−	1+2−	1+2−
Egocentric	2−3−	2−3−	2−2−
Egotistical	2−3−	2−3−	2−2−
Emotional	4−4−	4−2+	4−1+
Energetic	1+3+	1+3+	1+3+
Extraverted	1+3−	1+1+	1+3−
Generous	1+2+	2+1+	1+2+
Hard	2−4+	2−4+	2−3+
Humorous	1+2+	1+2+	1+2+
Humourless	1−5−	4+1−	1−5−
Impulsive	1+4−	3−4−	1+4−
Intelligent	5+4+	5+5+	5+5+
Intolerant	2−4−	2−4−	2−5−
Introverted	1−5+	1−1−	1−1−
Naive	5−4−	1−5−	5−3−
Orderly	3+3+	3+3+	3+2+
Punctual	3+3+	3+1−	3+3+
Rational	4+5+	3+5+	3+4+
Sarcastic	2−5+	2−3−	2−3−
Tolerant	2+1+	2+2+	2+5+

Note: 5+4+ means primary loading on the positive pole of Factor V, secondary loading on the positive pole of Factor IV. 1−1− means primary loading on the negative pole of Factor I, absence of secondary loading greater than 1/3.73 times the primary loading.

Testing a corollary of the lexical hypothesis

To investigate the extent to which the gaps that occur in the AB5C taxonomy are general across the three languages, we correlated the frequencies of traits with projections above 0.2 in each of the 90 AB5C facets among the three languages, using the joint structure. Between Dutch and English, the correlation was 0.58; between Dutch and German 0.33; and between English and German 0.48. Overall, these data lend modest support to this aspect of the lexical hypothesis. The pattern of correlations reflects the differences in selection procedures between the Dutch and German studies: due to the inclusion of capacity terms, the German Factor V attracts many adjectives, whereas the Dutch Factor V does not; the English set is in between. In regard to other discrepancies, it is hard to say whether slight differences in the positions of the axes, in the selections of traits (which would also influence the axis positions), or genuine differences between the languages are responsible. Finally, the cross-national correlations are attenuated by the limited numbers of traits; this limitation holds especially for the German set (430 traits, *versus* 551 in the Dutch set and 540 in the English set).

A specific mechanism that is at least partly responsible for the observed positive correlations relies in the distinction between two kinds of trait adjectives: concordant traits, whose primary and secondary loadings are both positive or both negative, and discordant traits, combining positive and negative loadings. As all factors are scored in the socially desirable direction, discordant traits (e.g., Dominant, which is in the

I + II − cell of Table 6) combine socially desirable and undesirable aspects. In that important sense, discordant traits are more ambiguous than concordant traits. This ambiguity may lead to lowered communalities, as different respondents have different interpretations and uses for such terms. Consequently, discordant cells are generally less likely to be well filled. The effect is illustrated in Table 6, which contains 43 incomplete cells (disregarding cells such as I + I −, which are empty by definition); of these, a majority (27) are discordant cells.

Similar findings may be observed in the Dutch (Hofstee and De Raad, 1991) and English (Hofstee *et al*., 1992) AB5C solutions. These findings suggest that the study of judgments of personality is on safer grounds with concordant items than with the present mixtures of concordant and discordant ones, although much subtlety would become lost upon discarding discordant descriptors: doing so would reduce the personality sphere from a generalized circumplex to a generalized bipolar simplex structure.

Additional analyses

So far, the overall conclusion of the present study seems to be that the Big-Five factors are replicable cross-nationally, but only weakly so. Haven and Ten Berge (1977) demonstrated that two factors are judged the same when their congruence coefficient is at least 0.85 (corresponding to an angle of about 30°). None of the factors from the original publications meets this criterion; even after rotation to joint cross-national position, fewer than half of the congruences are 0.85 or higher. We carried out additional analyses to ascertain that these disappointing results were not due to method artifacts or to unreliability of the three factor solutions.

To control for differences in data-analytic procedures among the three research teams, we used the three sets of raw data, rather than the published factor solutions. Cases with more than 10 per cent missing values were deleted, resulting in a reduction of the German sample from 802 to 770, and other missing values were replaced by the variable mean. In the Dutch data, the second subsample with self-ratings only was removed, retaining 400 subjects. All data were standardized row-wise. Principal component analysis and varimax rotation were applied to each of the data sets.

Congruence coefficients between corresponding factors are included in Table 3. Individual coefficients change somewhat, but the averages remain about the same. The procedure was repeated for just the sets of 126 adjectives that are common to the three data sets. In the present data, the congruences increased only a little (by about 0.015 on average). The reason why the increase is slight is in the selection of the 126 common terms: they were selected for their translatability, not for their fitness as factor markers in a particular language as in the imposed-etic approach.

To determine the impact of unreliability, two approaches were followed: (i) self-ratings and peer ratings in each of the three data sets were analysed separately, and congruence was assessed between the corresponding factors within each subset; (ii) within each data set, subjects were split at random into two subsets, and congruence was assessed between the corresponding factors from the random subsets. In both cases, congruences were taken over all items and over just the subset of 126 common items. As the results were virtually identical for these two variations, only the congruences over 126 items are reported in Table 7. These reliabilities serve as a

Table 6. Representative adjectives for the AB5C Facets in Dutch, English and German.

	I+	I−	II+	II−	III+	III−	IV+	IV−	V+	V−
I+	talkative		blijgeestig	baasachtig	wilskrachtig	onbesuisd	zelfverzekerd	heethoofdig	besluitvaardig	
	freiheraus		cheerful	domineering	alert	reckless		excitable	inventive	
			warmherzig	diktatorisch	zielsicher	leichtsinnig		emotional	ideenreich	
I−		gesloten	bescheiden	nors	degelijk	laks	bedaard	zenuwachtig	diepzinnig	gedwee
		introverted	modest	cold	careful	lazy	unemotional	fretful	introspective	unimaginative
		kontaktscheu	sanftmütig	kleinlich	sparsam	ehrgeizlos		selbstzweiflerisch		denkschwach
II+	opgewekt		mild		vlijtig	onnozel	kalm	gevoelig	ruimdenkend	volgzaam
	merry	shy	tolerant		responsible		unenvious	emotional	deep	simple
	kontaktfreudig	zurückhaltend	rücksichtsvoll		pflichtbewusst		gelassen	gefühlsbetont	schöpferisch	
II−	fel	stug			streng	onberekenbaar		opvliegend	ironisch	
	aggressive	unfriendly		harsh		inconsistent	insensitive	moody		
	dominant	eigenbrötlerisch		herrschbegierig	gestreng	arbeitsscheu	gefühlarm	launenhaft		
III+	energiek	gereserveerd	zorgzaam		zorgvuldig		evenwichtig			braaf
	energetic	reserved	cooperative		organized				knowledgeable	
	tatkräftig	unspontan	aufrichtig	streng	arbeitsam		gefestigt		klug	amusisch
III−	ongeremd	passief		egocentrisch		nonchalant		instabiel	vrijgevochten	lichtgelovig
		unergetic		inconsiderate		disorganized				
	heissblütig	wachsweich	musseliebend	rücksichtslos		flatterhaft	aalglatt	labil	kunstsinnig	unkundig
IV+	optimistisch	afstandelijk	verdraagzaam	glashard	consequent	gemakzuchtig			onafhankelijk	oppervlakkig
	confident	uncommunicative	undemanding	unsympathetic	thorough	sloppy			brilliant	unintellectual
	selbstvertrauend	bedächtig	geduldig	habsüchtig	strebsam	korrupt			kenntnisreich	dumm
IV−	explosief	onzeker	zacht	driftig	ijverig	wispelturig		emotioneel		moederlijk
		self-pitying	sympathetic	irritable		scatterbrained			complex	
	temperamentvoll	ängstlich	menschlich	querköpfig		unbeständig		verletzbar	phantasievoll	unlogisch
V+	ondernemend	introvert	sympathiek	sarcastisch	vastberaden	rebels	rationeel		kritisch	
	magnetic				exacting				creative	
	dynamisch		tolerant	anspruchsvoll	ehrgeizig		superklug	sensibel	geistvoll	
V−	praatziek	bedeesd	gewillig	hebberig	gedisciplineerd			paniekerig		
		bland		scornful		haphazard		envious		uncreative
		mundfaul	gutgläubig	halsstarrig		vergesslich				unbegabt

proper baseline for the cross-national congruences, which are necessarily taken over just the 126 common items.

Table 7 shows that the within-language congruences (mean, 0.89) are much higher than the cross-language coefficients (mean before rotation, 0.70; after rotation, 0.76). Moreover, the reliabilities are underestimates because the within-language congruences apply to factors that are derived from half-size (in the Dutch case, one-third-size) samples. Thus the modest cross-language congruences cannot be explained solely by unreliability of the factor structures.

DISCUSSION

One conclusion from this study is that the Big-Five factors of personality recur in a relative sense across the present sample of Indo-European languages. In the framework of the present emic approach—as opposed to the imposed-etic strategy by which highly selected marker items are translated into other languages—strict correspondence in the sense of congruence coefficients averaging above 0.85 is not to be found (see also De Raad *et al.*, 1996). A further reservation should be made with respect to Factor V. This factor replicates only if ability terms are included, as was the case in the German and English trait sets.

Another finding is that systematic discrepancies in the positions of the factors are to be expected across languages. If the world conformed to simple structure, factors would be stable and easily interpretable. In the absence of simple structure, however, factor positions shift easily, and impressions of global correspondence obscure subtle but systematic differences in angular positions. The multiple-circumplex reconstruction of the data by the AB5C model captures these discrepancies.

It should be clear from the results that none of the published solutions can claim to give the natural positions of the Big Five (see also De Raad *et al.*, 1996). Maybe such natural positions exist, and can ultimately be found; maybe they do not. Clearly, however, the present factor-pure (I+I+ to V−V−) clusters, or any other sets of traits, do not occupy privileged interpretive positions. The same goes for the

Table 7. Within-language analyses: the reliability of factor structures in Dutch, English, and German

	I	II	III	IV	V	Mean
			Dutch Factors			
Self *versus* peer	0.91	0.86	0.87	0.88	0.78	0.86
Random split	0.90	0.88	0.85	0.82	0.73	0.84
			English Factors			
Self *versus* peer	0.93	0.92	0.91	0.77	0.90	0.89
Random split	0.92	0.92	0.93	0.79	0.87	0.89
			German Factors			
Self *versus* peer	0.97	0.96	0.92	0.92	0.94	0.94
Random Split	0.96	0.97	0.92	0.94	0.92	0.94

90 AB5C cells in Table 6. A different rotation of the axes would draw together somewhat different sets of trait terms, as can be verified by comparing the national solutions (Hofstee and De Raad, 1991; Hofstee *et al.*, 1992) with the present structure. All the AB5C model does is to provide a finer grid for keeping track of these shifts.

Apart from the instability of factors, this study emphasizes the indeterminacy of single trait terms. In everyday use, the meaning of trait adjectives is coloured by the context—the person being discussed, the common history of the discussants, the script of the discussion, and so forth. Context-free scientific applications need other aids to fix their meaning. No single trait term can function as a prototype providing an Archimedic point upon which meaning can be pinned, as each adjective has a large error component resulting from divergent idiosyncratic uses, in addition to its modest common component. Consequently, labelling factors in terms of substantives derived from trait adjectives (Extraversion, Agreeableness, Conscientiousness, Stability, Intellect) is to some extent arbitrary. Instead of making such arbitrary choices, a psychometric theory of scientific meaning is implemented here. It consists of averaging out specific and error components by clustering trait terms, and of recursive definition by contrasting facets with opposite and adjacent clusters in the trait space.

Studying the trait sphere is somewhat like looking through a kaleidoscope. Each slight rotation produces a different but equally fascinating configuration, characterized by symmetry and redundancy. With trait terms, however, we do not know precisely what the mechanism is. Until we do, some tolerance for indeterminacy is in order.

ACKNOWLEDGEMENT

Work by the fourth author on this project was supported by grant MH-49227 from the National Institute of Public Health, U.S. Public Health Service.

REFERENCES

Berry, J. W. (1969). 'On cross-cultural comparability', *International Journal of Psychology*, **4**: 119–128.

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

De Raad, B., Perugini, M. and Szirmák, Z. (1997). 'In pursuit of a cross-lingual reference structure of personality traits: comparisons between five languages', submitted to *European Journal of Personality*.

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

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

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

Hakstian, A. R. (1973). 'Procedures for the factor analytic treatment of measures obtained on different occasions', *British Journal of Mathematical and Statistical Psychology*, **26**: 219–239.
Haven, S. and Ten Berge, J. M. F. (1977). 'Tucker's coefficient of congruence as a measure of factorial invariance: an empirical study', unpublished manuscript, University of Groningen.
Hofstee, W. K. B. (1990). 'The use of everyday personality language for scientific purposes', *European Journal of Personality*, **4**: 77–88.
Hofstee, W. K. B. and De Raad, B. (1991). 'Persoonlijkheidsstructuur: de AB5C taxonomie van Nederlandse eigenschapstermen' ['Personality structure: The AB5C taxonomy of Dutch trait terms'], *Nederlands Tijdschrift voor de Psychologie*, **46**: 262–274.
Hofstee, W. K. B., De Raad, B. and Goldberg, L.R. (1992). 'Integration of the Big Five and circumplex approaches to trait structure', *Journal of Personality and Social Psychology*, **63**: 146–163.
John, O. P. (1990). 'The "Big-Five" factor taxonomy: dimensions of personality in the natural language and in questionnaires'. In: Pervin, L.A. (Ed), *Handbook of Personality: Theory and Research*, pp. 66–100, Guilford, New York.
Katigbak, M.S., Church, A. T. and Akamine, T. X. (1996). 'Cross-cultural generalizability of personality dimensions: relating indigenous and imported dimensions in two cultures', *Journal of Personality and Social Psychology*, **70**: 99–114.
Neisser, U., Boodoo, G., Bouchard, T. J. Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D.F., Loehlin, J. C., Perloff, R., Sternberg, R. J. and Urbina, S. (1996). 'Intelligence: knowns and unknowns', *American Psychologist*, **51**: 77–101.
Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur: zur Validität des Fünf-Faktoren-Modells der Persönlichkeit* [Language and Personality Structure: on the Validity of the Five-Factor Model of Personality], Roderer, Regensburg, Germany.
Ostendorf, F. and Angleitner, A. (1994). 'The five-factor taxonomy: robust dimensions of personality description', *Psychologica Belgica*, **34**: 175–194.
Peabody, D. (1984). 'Personality dimensions through trait inferences', *Journal of Personality and Social Psychology*, **46**: 384–403.
Peabody, D. and Goldberg, L. R. (1989). 'Some determinants of factor structures from personality-trait descriptors', *Journal of Personality and Social Psychology*, **57**: 552–567.
Rodale, J. I. (Ed). (1961). *Synonym Finder*, Rodale Books, Emmaus, PA.
Ten Berge, J. M. F. (1977). 'Orthogonal Procrustes rotation for two or more matrices', *Psychometrika*, **42**: 267–276.
Tucker, L. J. (1951). 'A method for synthesis of factor analysis studies', Personnel research section report No. 984, Department of the Army, Washington, DC.