REPLY TO EYSENCK*

PAUL T. COSTA JR and ROBERT R. McCRAE

Gerontology Research Center, National Institute on Aging, NIH Baltimore, MD, U.S.A.

Summary—In this response to Eysenck's comments we argue that a contemporary review of the literature would favor the five-factor model; we attempt to explain the observed correlations between scales that measure different factors; and we reiterate our view that the systematic description of personality must precede, not follow, personality theory.

In the interest of brevity, we might limit our reply to Eysenck (1992) to an invitation to the reader to re-read our original article (Costa & McCrae, 1992); most of the points he makes are anticipated and answered there. For example, he discusses the Zuckerman, Kuhlman, Thornquist and Kiers (1991) article and comments that their five factors do not match ours very well. We acknowledge this, and showed how a re-rotation could substantially improve the match. He points out that cross-observer correlations for Conscientiousness (C) are "derisorily small," noting a 0.30 between peer raters. But Table 2, from which he extracts this figure, also shows peer/self correlations of 0.46 for C—higher than those found for the uncontested dimensions of Neuroticism (N) (0.33) and Extraversion (E) (0.43). There are some points, however, we would like to clarify here.

In our original article, we argued that the dimensions of the five-factor model are real, pervasive, universal, and biologically based. In his reply, Eysenck grants these points and acknowledges that they are necessary features of basic dimensions of personality, but argues that they are not sufficient: they do not distinguish the five-factor model from its competitors—at least from Eysenck's three-factor model. He is of course correct in one sense. If we assume that the three-factor model is a subset of the five-factor model [Psychoticism (P) being viewed as a combination of low Agreeableness (A) and low C], then the dimensions of the three-factor model must also be real, pervasive, universal, and biologically based. We are happy to grant this.

The three-factor model, however, is not comprehensive, most obviously because it does not account for traits related to Openness (O). By listing a catalog of related traits, we intended to show that this dimension has recurred in too many conceptualizations to be ignored. By reviewing its minimal correlation with IQ, we have shown that O cannot be reduced to intellectual ability. At one time, Eysenck suggested that O "represents possibly the opposite end of a continuum to psychoticism" (Eysenck & Eysenck, 1985, p. 138), but in fact the correlation between O and P is 0.05, $N = 586$, NS (McCrae & Costa, 1985). It would seem that, at a minimum, Eysenck should recognize the need for a four-factor model: N, E, P and O.

We concur with Eysenck's view that purely psychometric criteria cannot resolve the issue of which factors are basic (although Everett's 1983 formalization of the criterion of factor replicability is a marked advance over eigenvalue rules). Ultimately, the decision of what factor structure is best requires scientific judgment on the recurrence of regular patterns that make sense of the broadest range of data.

Eysenck claims that the five-factor model has not been recovered by researchers "outside the narrow circle of five-factor enthusiasts," and cites a 1983 review by Royce and Powell as evidence of the wider acceptability of the three-factor model. It is possible that, as of 1983, there was more evidence in favor of a three-factor model than a five-factor model, but much has changed since then. Perhaps we should set aside the voluminous research on trait adjectives (e.g. Goldberg, 1989) and adjective-based scales (e.g. Trapnell & Wiggins, 1990; Brand & Egan, 1989), because the five factors were first discovered in analyses of adjectives, and these studies could be considered

---

replications instead of independent rediscoveries. We could also set aside research using Norman's (1963) rating scales (e.g. Yang & Bond, 1990) because these scales were based on analyses of trait adjectives, and studies which employed the NEO Personality Inventory (Costa & McCrae, 1985; McCrae, 1989) because it was developed explicitly to measure the five-factor model. A contemporary meta-analysis would still have to consider Amelang and Borkenau (1982), Krug and Johns (1986), Conn and Ramanaiah (1990), Noller, Law and Comrey (1987, especially as reanalyzed by Boyle, 1989), Zuckerman, Kuhlman, Thornquist and Kiers (1991, especially as reanalyzed in our article), Loehlin (1987), Botwin and Buss (1989), and Lanning and Gough (1991)—all studies in which most or all of the five factors were identified in analyses of a wide variety of instruments that had originated in other contexts. Like Wiggins and Trapnell (in press), we believe the weight of the evidence has now clearly shifted in favor of the five-factor model.

INTERCORRELATIONS AMONG THE FACTORS

One point of Eysenck's commentary is likely to cause confusion. Early in his article he suggests that C is not a separate factor; instead, it is a rather small part of P, as illustrated in his Fig. 1(a). Later, he notes the sizeable correlation (−0.49) between C and N domain scales of the NEO-PI-R, and suggests that traits related to C can be explained theoretically as aspects of N. Perhaps he intended to say that C is a blend of low P and low N (just as we have argued that P is a blend of low C and low A).

More generally, his point is that the five do not really represent the highest order, and thus the most basic, factors. He is not alone in this view: Digman (1991), a well-known advocate of the five-factor model, has also noted correlations between scales measuring the five factors and suggested that there may be two higher-order factors, one contrasting N with A and C, the other combining E and O. He has argued that the first superfactor might represent Socialization, the second, Self-actualization.

We do not share this view; we think that the five dimensions themselves are essentially orthogonal and irreducible. How, then, do we account for the fact that scales measuring the five factors are often intercorrelated? There are two explanations. First, biases in implicative meaning can create spurious correlations. The factor that Digman interprets as Socialization might equally well be characterized as evaluation: A and C are desirable characteristics, whereas N is undesirable. Individuals who view themselves positively will tend to rate themselves as somewhat higher in A and C and lower in N than they actually are. This bias is not necessarily shared by other observers, and that may account for the fact that orthogonal factor scores have higher cross-observer validity than do (spuriously) oblique factor scores (McCrae & Costa, 1989).

The other reason is substantive. As we noted in our original paper, personality traits are not neatly clustered into discrete domains. Many, perhaps most, traits overlap two or more of the separate factors; for example, interpersonal warmth is related to both E and A. If an instrument samples traits broadly, it is likely to include many such mixed traits, and scales based on unit weighting (like the domain scales of the NEO-PI-R) will probably show intercorrelations of some kind.*

If these intercorrelations were consistent across instruments—if many different samplings of traits led to the same pattern of associations among domains—we might well conclude that the five factors were themselves oblique, correlated in ways that could be meaningfully summarized by higher-order factors. But they are not.

Table 1 shows second-order analyses of three instruments intended to measure the five-factor model: a set of adjective scales developed by Goldberg (1989) and administered in a transparent format, that is, with similar items grouped together; another set of adjective scales developed by Trapnell and Wiggins (1990) to extend their earlier work on the interpersonal circumplex; and the NEO-PI-R. All three have shown convergent and discriminant validity as measures of the five factors. The data are taken from peer ratings of participants in the Baltimore Longitudinal Study

---

*Where uncorrelated measures of the factors are needed, we recommend orthogonal factor scores (cf. Table 2 in our original article). These are calculated automatically in NEO-PI-R computer-administered testing, and scoring weights from the standardization sample are provided in the NEO-PI-R manual.

Table 1. Second-order factor analyses of three measures of the five-factor model

Scale	Two-factor solution		Three-factor solution		
	I	II	I	II	III
Transparent Trait Rating Form[a]					
Low Emotional Stability (N)	**0.92**	−0.19	**0.89**	−0.03	−0.29
Extraversion (E)	−0.10	**0.78**	−0.17	**0.94**	0.14
Intellect (O)	−0.29	**0.81**	−0.23	**0.46**	**0.71**
Agreeableness (A)	**−0.84**	0.29	**−0.87**	0.31	0.14
Conscientiousness (C)	−0.35	**0.63**	−0.21	0.02	**0.91**
Revised Interpersonal Adjective Scales—Big Five Version[b]					
Neuroticism (N)	**0.79**	−0.11	**0.90**	−0.15	0.02
Dominance (E)	0.25	**0.81**	0.10	**0.91**	−0.01
Openness (O)	**−0.44**	**0.49**	−0.11	0.08	**0.95**
Love (A)	**−0.87**	−0.01	**−0.81**	−0.14	0.30
Conscientiousness (C)	−0.33	**0.67**	−0.30	**0.55**	**0.41**
Revised NEO Personality Inventory[c]					
Neuroticism (N)	**0.85**	−0.10	**0.59**	−0.09	**−0.61**
Extraversion (E)	−0.15	**0.85**	0.03	**0.88**	0.23
Openness (O)	−0.06	**0.90**	−0.21	**0.88**	−0.10
Agreeableness (A)	**−0.66**	0.18	**−0.94**	0.10	0.06
Conscientiousness (C)	**−0.72**	0.00	−0.05	0.06	**0.92**

These are varimax-rotated principal components from single peer ratings; see Costa and McCrae (in press) for details. Loadings greater than 0.40 are given in **boldface**.
[a] Goldberg (1989) $N = 137$.
[b] Trapnell and Wiggins (1990) $N = 156$.
[c] Costa, McCrae and Dye (1991) $N = 227$.

of Aging (Costa & McCrae, in press); factors have been rearranged and reflected to simplify comparisons.

The first two columns show a two-factor solution. Digman (1991) predicted a Socialization factor of N versus A and C and a Self-actualization factor of E and O. This pattern is indeed found in the NEO-PI-R data, but not in the other two instruments. In the two adjective instruments, the first factor is N versus A; C joins E and O in the second factor. Digman's two-factor solution is apparently not replicable across instruments.

The last three columns show a three-factor solution that might be consistent with Eysenck's model. Presumably he would hypothesize N, E, and P factors, with P defined chiefly by low A. From his commentary, it appears that he would predict that C should show negative loadings on both the N and P factors and that O would join with E. But none of the three instruments shows this pattern. The first factor in each is N versus A; the second shows E twice with O, but once with C. The third factor is C with O, O with C, or C versus N. Readers who wish to analyze the self-report data in Costa, McCrae and Dye's (1991) Table 5 will find yet another three-factor solution: C and E versus N, E and O, and A alone. Nowhere in any of the analyses is a P-like factor defined solely by low A with low C to be found.

In short, two-factor solutions are not replicable. Three-factor solutions are not replicable. Only five-factor solutions have been replicated time after time, in diverse instruments (cf. Borkenau & Ostendorf, 1990). That is why five factors are basic.

THE ROLE OF THEORY

In our article, we rejected the idea that personality structure should be determined by theory, noting particularly that our knowledge of neuroscience is too primitive to guide decisions about personality description. Predictably, Eysenck rises to the defense of biological theories of personality. We share with him the view that progress is being made in these areas, and we fully support research on personality and psychobiology. We remain skeptical, however, that sufficient progress has been made to allow much contribution to the issue of basic dimensions of personality.

In our view, outside a small circle of Eysenckian enthusiasts, there is little support for any particular psychobiological theory of personality. Block (1977) summarized criticisms of the P scale and its theoretical relation to psychosis and psychopathy. A recent chapter by Amelang and Ullwer (1991) reported an ambitious attempt to replicate findings on the psychophysiological, perceptual, learning, and motor correlates of E in a German sample. The authors concluded that "none of the hypotheses concerning the excitation–inhibition-balance could be confirmed" (p. 310). Gray (1991), who has devoted a distinguished career to the study of personality and neuropsychology,

freely admitted that "human individual differences ... are still difficult to link to the workings of the brain" (p. 126). Surely this is the area in which "premature crystallizations of spurious orthodoxy" are to be feared.

Eysenck's ultimate goal is the creation of a paradigm for personality psychology that can guide fruitful normal science. We concur in the need for such a paradigm, but believe that at this point in the development of our science what is needed is not theory, but a systematic method of description. There has been more than enough description of individual differences, but little coherent system. The paradigm that has brought most order to this field is the one which seeks the broadest possible themes, the most common and fundamental dimensions, and only after exhausting their possibilities moves on to add new dimensions of comparable breadth and significance. This approach, of course, is precisely that adopted by Eysenck in his ground-breaking emphasis on E and N, and later P (Costa & McCrae, 1986). We regard the five-factor model not as a competing paradigm, but as the result of normal science within Eysenck's own descriptive paradigm.

REFERENCES

Amelang, M. & Borkenau, P. (1982). Über die faktorielle Struktur und externe Validität einiger Fragebogen-Skalen zur Erfassung von Dimensionen der Extraversion und emotionalen Labilität [On the factor structure and external validity of some questionnaire scales measuring dimensions of extraversion and neuroticism]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 3,* 119–146.

Amelang, M. & Ullwer, U. (1991). Correlations between psychometric measures and psychophysiological as well as experimental variables in studies on extraversion and neuroticism. In Strelau, J. & Angleitner, A. (Eds), *Explorations in temperament: International perspectives of theory and measurement* (pp. 287–316). New York: Plenum.

Block, J. (1977). P scale and psychosis: Continued concerns. *Journal of Abnormal Psychology, 86,* 431–434.

Borkenau, P. & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences, 11,* 515–524.

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

Boyle, G. J. (1989). Re-examination of the major personality-type factors in the Cattell, Comrey and Eysenck scales: Were the factor solutions by Noller *et al.* optimal? *Personality and Individual Differences, 10,* 1289–1299.

Brand, C. R. & Egan, V. (1989). The 'Big Five' dimensions of personality? Evidence from ipsative, adjectival self-attributions. *Personality and Individual Differences, 10,* 1165–1171.

Conn, S. R. & Ramanaiah, N. V. (1990). Factor structure of the Comrey Personality Scales, the Personality Research Form-E, and the five-factor model. *Psychological Reports, 67,* 627–632.

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

Costa, P. T. Jr & McCrae, R. R. (1986). Major contributions to personality psychology. In Modgil, S. & Modgil, C. (Eds), *Hans Eysenck: Consensus and controversy* (pp. 63–72, 86, 87). Barcombe, Lewes, Sussex, England: Falmer.

Costa, P. T. Jr & McCrae, R. R. (in press). Trait psychology comes of age. In Sonderegger, T. B. (Ed.), *Nebraska symposium on motivation: Psychology and aging.* Lincoln, NE: University of Nebraska Press.

Costa, P. T. Jr & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13(6),* 653–665.

Costa, P. T. Jr, McCrae, R. R. & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12,* 887–898.

Digman, J. M. (August, 1991). The Big Five: Up, down, and beyond. In Strack, S. (Chair), *Beyond the big five.* Symposium presented at the *American Psychological Association Convention,* San Francisco.

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

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

Eysenck, H. J. & Eysenck, M. (1985). *Personality and individual differences.* London: Plenum.

Goldberg, L. R. (June, 1989). Standard markers of the Big Five factor structure. Paper presented at the *First International Workshop on Personality Language,* Groningen, The Netherlands.

Gray, J. A. (1991). The neuropsychology of temperament. In Strelau, J. & Angleitner, A. (Eds), *Explorations in temperament: International perspectives on theory and measurement* (pp. 105–128). New York: Plenum.

Krug, S. E. & Johns, E. F. (1986). A large scale cross-validation of the second-order personality structure defined by the 16PF. *Psychological Reports, 59,* 683–693.

Lanning, K. & Gough, H. G. (1991). Shared variance in the California Psychological Inventory and the California Q-Set. *Journal of Personality and Social Psychology, 60,* 596–606.

Loehlin, J. C. (1987). Heredity, environment, and the structure of the California Psychological Inventory. *Multivariate Behavioral Research, 22,* 137–148.

McCrae, R. R. (1989). Why I advocate the five-factor model: Joint analyses of the NEO-PI and other instruments. In Buss, D. M. & Cantor, N. (Eds), *Personality psychology: Recent trends and emerging directions* (pp. 237–245). New York: Springer-Verlag.

McCrae, R. R. & Costa, P. T. Jr (1985). Comparison of EPI and Psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences, 6,* 587–597.

McCrae, R. R. & Costa, P. T. Jr (1989). Different points of view: Self-reports and ratings in the assessment of personality. In Forgas, J. P. & Innes, M. J. (Eds), *Recent advances in social psychology: An international perspective* (pp. 429–439). Amsterdam: Elsevier Science Publishers.

Noller, P., Law, H. & Comrey, A. L. (1987). Cattell, Comrey, and Eysenck personality factors compared: More evidence for the five robust factors? *Journal of Personality and Social Psychology, 53,* 775–782.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66,* 574–583.

Royce, J. R. & Powell, A. (1983). *Theory of personality and individual differences: Factors, systems, and processes.* Englewood Cliffs, NJ: Prentice-Hall.

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

Wiggins, J. S. & Trapnell, P. D. (in press). Personality structure: The return of the Big Five. In Briggs, S. R., Hogan, R. & Jones, W. H. (Eds), *Handbook of personality psychology.* New York: Academic Press.

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

Zuckerman, M., Kuhlman, D. M., Thornquist, M. & Kiers, H. (1991). Five (or three) robust questionnaire scale factors of personality without culture. *Personality and Individual Differences, 12,* 929–941.