PERGAMON

Personality and Individual Differences 30 (2001) 1147–1160

PERSONALITY AND INDIVIDUAL DIFFERENCES

www.elsevier.com/locate/paid

Age changes in personality traits and their heritabilities during the adult years: evidence from Australian twin registry samples

J.C. Loehlin[a,*], N.G. Martin[b]

[a]University of Texas, Department of Psychology, Austin, TX 78712, USA
[b]Queensland Institute of Medical Research, Brisbane, Australia

Received 19 November 1999; received in revised form 10 March 2000

Abstract

Short versions of four Eysenck personality scales had been included in questionnaires given to several adult samples from the Australian Twin Registry, comprising altogether some 5400 pairs. Means and regressions with age are compared for three samples at average ages of 23, 37, and 61 years, and for two samples of retested individuals, one tested twice at average ages of 29 and 37 years, and one tested three times at average ages of 48, 56, and 62 years. For both males and females the trends for Psychoticism (P), Extraversion (E), and Neuroticism (N) were generally downward with age, and for Lie (L), upward. However, in the longitudinal sample between ages 56 and 62 the trends for P, E, and L stopped or reversed, although N continued downward. Heritabilities were reasonably stable across age for P, E, and N, and the effects of shared environments negligible, but L showed some influence of shared environment as well as genes in all but the oldest age group. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Eysenck scales; Heritability; Age trends; Twins

1. Introduction

Age trends for the four Eysenck scales Psychoticism (P), Extraversion (E), Neuroticism (N), and Lie (L) were presented by Martin and Jardine (1986), based on data from short versions of these scales given to a large sample of adult Australian twins. They found a tendency toward decrease in age with scores on P, E, and N, and an increase in L. Much more information is now available from this population. There are now several twin samples spanning a substantial age

* Corresponding author. Tel.: +1-512-475-7008; fax: +1-512-471-6175.
E-mail address: loehlin@psy.utexas.edu (J.C. Loehlin).

PII: S0191-8869(00)00099-4

range from an average of about 23 years of age in a young adult twin sample (range 17–29 years) to an average of 61 years of age in a sample of older twins (range 50–92 years). Most of these received versions of the Eysenck questionnaire scales. However, the scales used differed somewhat from sample to sample, and there has been no systematic comparison of age trends across them. We undertake this here, using scales based on identical questionnaire items in three samples — the younger and older twins mentioned above, and a third sample intermediate in age although overlapping with both at the extremes (age range 24–92).

The data also permit comparisons based on the same individuals measured on Eysenck scales on more than one occasion. For one fairly large group of twins, data are available from two testings roughly 8 years apart in early adulthood. For a smaller group, data are available on three occasions spanning some 14 years later in life. Thus estimates can also be made based on the same persons at different ages.

Besides simple age trends, the fact that these individuals are twins also permits a test for possible age trends in heritabilities during the adult years. McCartney, Harris, and Bernieri entitled their 1990 meta-analysis of age trends in twin studies 'Growing up and growing apart'. This title reflected a tendency for twin pairs as they grew older to become more different from each another in abilities and personality. For personality traits, the tendency to grow apart was about the same for monozygotic (MZ) and dizygotic (DZ) twin pairs, suggesting that the heritability of personality remained roughly constant, while the influence of shared environment decreased. The ages of the twins in these studies ranged from 1 to 50 years, but the majority were relatively young — the median age overall was about 16. Thus whether these trends would characterize the adult years remains in some doubt. A study of Australian twins aged 12–18 found quite different age trends than those noted for adult twins above: no change in means with age for P and N, a decrease in L, and an increase in E for females only (Macaskill, Hopper, White & Hill, 1994).

For two traits, Neuroticism and Extraversion, it will be possible to make comparisons with a large Finnish study in which twins aged 18–59 were tested twice 6 years apart using mail questionnaires (Viken, Rose, Kaprio & Koskenvuo, 1994).

2. Method

2.1. Participants

The twins were volunteer respondents drawn from the Australian Twin Registry, all of whom participated via mail questionnaires. The intermediate-age group (mean age 37 after eliminating overlap — see below) was a sample drawn in 1988 as a follow-up to an earlier 1980 study. About 2700 pairs completed a long questionnaire on health and lifestyle matters that included a number of personality items. Among these were items constituting short 12-item versions of Eysenck's Extraversion (E), Neuroticism (N), Psychoticism (P), and Lie (L) scales (Eysenck, Eysenck & Barrett, 1985). Those participating in both 1980 and 1988 provide the two-occasion data.

The young twin sample (mean age 23) was assessed in 1990 from a group of twins who had been too young to participate in the original 1980 study. Again, items constituting the four 12-item Eysenck scales were embedded within a longer questionnaire. Responses on these items were available for about 1500 pairs in this group.

The third, older sample (mean age 61) was assessed in 1993 — again, with Eysenck items as part of a longer questionnaire. These included the same 12 items as in the other two samples for the E, N, and P scales, and 11 of the 12 items for the L scale (the inadvertently omitted item was item 40 in Eysenck et al., 1985, Appendix 2). About 1200 pairs responded to the questionnaire items in this study. About one-third of these had also participated as members of the intermediate-aged sample and the original sample of which it was a follow-up, providing the three-occasion sample.

Zygosity was diagnosed by similarity of appearance and mistaken identity items on the questionnaire, a procedure that typically provides agreement of 95% or better with the results of blood testing. A subsample of the twins in the studies reported here took part in a parallel study of asthma in which 198 same-sex pairs were typed for 11 independent highly polymorphic markers. No errors in the previous questionnaire-based diagnosis of zygosity were detected (Duffy, 1994).

In all three samples, there were more MZ than DZ twins, and more females than males, as is typical in studies with volunteer twins (Lykken, Tellegen & DeRubeis, 1978).

Table 1 summarizes the samples. To avoid duplication, individuals participating in both the middle and old samples are included only once, with the old sample (the young and middle samples were non-overlapping). Similarly, for the retested samples, individuals in both the twice- and thrice-tested samples are included only once, with the latter. However, individuals in the longitudinal samples are included in the cross-sectional samples.

Note that the male and female subsamples, although different in size, are closely similar in mean and range of ages, and that the cross-sectional and longitudinal samples both span a comparable and substantial adult age range.

2.2. Measures

As mentioned, items constituting short versions of the Eysenck scales were included within each questionnaire, although they were differently ordered and in different item contexts in each. Thus it was possible to provide scale scores based on identical items — 12 each for Extraversion,

Table 1
Summary of samples

Sample	Males			Females			Years tested
	Mean age	Range	No.	Mean age	Range	No.	
Separate							
Young	23.2	18–29	1464	23.3	17–29	2265	1990–1992
Middle	37.2	24–87	1704	37.8	24–87	2952	1988–1990
Old	61.4	50–90	866	61.2	50–92	2091	1993–1995
Retested — 2 times							
1st	28.9	18–77	1683	29.6	18–78	2902	1980–1982
2nd	37.2	24–86	1683	37.8	24–87	2902	1988–1990
Retested — 3 times							
1st	47.6	36–72	327	47.5	36–77	914	1980–1982
2nd	55.7	44–81	327	55.7	44–85	914	1988–1990
3rd	61.9	50–87	327	61.9	50–92	914	1994–1995

Neuroticism, and Psychoticism, and 11 for Lie — in each of the three main samples. For the retested samples, the same E, N, and L scales were available, and an eight-item subset of the P scale. The score in each case was taken simply as the number of items answered in the keyed direction for the scale; however, if the respondent omitted four or more keyed items on a given scale, the scale was scored as missing. All scoring was done by computer program from item-level data files.

3. Results

3.1. Cross-sectional comparisons of means and age regressions

Table 2 shows means and standard deviations in the three samples, over all individuals for whom scores were available, including cases where only one of a pair responded. Males and females are shown separately.

In the case of the means, trends over the three age samples are clear and similar in males and females. Scores on P, E, and N tended to decrease with age, L to increase. The greatest amount of change occurred between the young and middle groups, with the trend from middle to old in the same direction but weaker. For the variabilities, little, if any, overall trend is present, except for P. On this highly skewed scale means and variances are correlated; as the scores decrease towards their lower limit of zero, the variability is restricted as well. A square-root transformation of P, $Pq = \sqrt{(P + 1)}$, decreases this dependence of variance on mean, and will be used for purposes of model fitting and significance testing.

The means themselves differ by sex: males have higher average scores on P and females average higher on E, N, and L; nevertheless the trends over age are quite similar for both sexes.

Table 3 suggests that on the whole the same relationships hold within the samples as between them — i.e., negative within-sample correlations with age for P, E, and N, and positive for L. The sizes of the correlations are small — a person's age is a weak predictor of his or her test score. The correlations are smallest for the young group, which is the most severely restricted in age range.

Table 2
Male and female means and standard deviations on Eysenck scales in Australian twin samples differing in age[a]

Sample	Means				SDs				No.
	P	E	N	L	P	E	N	L	
Males									
Young	3.71	7.66	4.43	3.89	1.90	3.46	3.15	2.45	1455
Middle	2.29	6.89	3.82	4.83	1.64	3.56	3.15	2.62	1697
Old	1.91	6.44	3.38	5.04	1.48	3.48	3.08	2.54	860
Females									
Young	2.84	8.04	5.67	4.50	1.73	3.37	3.14	2.42	2256
Middle	1.59	6.86	5.00	5.46	1.41	3.53	3.16	2.58	2937
Old	1.61	6.60	4.21	5.83	1.34	3.40	3.17	2.49	2069

[a] P = Psychoticism, E = Extraversion, N = Neuroticism, L = Lie. No. = average number of individuals (number varies slightly from scale to scale due to missing scores).

The general trends visible in Tables 2 and 3 were confirmed by fitting structural equation models to the means and covariances in the six age-sex groups. Because the four Eysenck dimensions are theoretically independent of one another, we have analyzed them as such. However, as is often the case, a number of the empirical correlations among the Eysenck scales in fact depart from zero, although modestly so. The largest is a negative correlation between Extraversion and Neuroticism. In the six age-sex samples this correlation ranged from −0.16 to −0.26, averaging −0.21. The next largest was between N and L, ranging from −0.08 to −0.17, averaging −0.13. Other correlations, weaker than these, but consistent in direction in at least five of the six groups, were P with E (positive), and N with P, L with E, and L with P (negative). Some of these correlations may reflect generalized tendencies toward positive or negative self-evaluation in subjects' responses, or generally conforming or nonconforming behavior. However, they do not imply a degree of redundancy among the scales that would necessitate a multivariate approach.

Table 4 shows the results of model fitting carried out for each of the four Eysenck scales.

For all of the scales except P, a simple model fit the data reasonably well. This model assumes constant regression of the trait on age in the six groups, a constant residual variance, and a constant

Table 3
Male and female within-group correlations with age in three samples[a]

Sample	P	E	N	L	No.
Males					
Young	0.02	0.01	−0.01	0.07	1455
Middle	−0.17	−0.12	−0.03	0.28	1697
Old	−0.08	−0.02	−0.06	0.25	860
Females					
Young	−0.05	−0.07	−0.00	0.05	2256
Middle	−0.03	−0.12	−0.06	0.23	2937
Old	−0.01	−0.06	−0.07	0.27	2069

[a] P = Psychoticism, E = Extraversion, N = Neuroticism, L = Lie. No. = average number of individuals (number varies slightly from scale to scale due to missing scores).

Table 4
Model fitting to means and covariances of age and personality scales in males and females at three adult ages[a]

Test	Pq	E	N	L
Fit of model				
χ^2, 18df	97.46*	25.60	28.66	39.73*
RMSEA	0.048	0.014	0.018	0.025
UL	0.058	0.027	0.029	0.036
Means equal for males and females? χ^2_{diff}, 1df	430.29*	5.72*	163.83*	177.86*
Means equal in three age groups? χ^2_{diff}, 2df	648.87*	50.20*	22.52*	131.62*
No age-trait covariance? χ^2_{diff}, 1df	32.46*	74.55*	6.83*	503.84*
Standardized age regression	−0.05	−0.08	−0.07	0.21

[a] Model for each trait assumes constant regression of trait on age, constant residual variance, and a constant sex difference in means. Pq = square-root-transformed P. RMSEA = root mean square error of approximation. UL = upper limit of 90% confidence interval for RMSEA. *$p < 0.05$.

sex difference in means at the three ages. With these large samples, some of the overall chi squares imply significant departures from perfect fit, but the descriptive index RMSEA (root mean square error of approximation; see Steiger, 1989; Browne & Cudeck, 1993) is well below the level of 0.05 generally considered indicative of a good fit. An advantage of RMSEA over some other fit indices is that confidence intervals are available for it; they permit us to reject for these traits with at least 90% confidence the hypothesis of a poor fit (RMSEA > 0.10) in the population.

For untransformed P the fit of the simple model was poor ($\chi^2_{18} = 416.97$, RMSEA = 0.11), reflecting the failure of the assumption of a constant variance across groups. The use of the square-root-transformed version (the Pq shown in the table) improved matters, yielding an acceptable RMSEA (0.048). The upper confidence limit of 0.058 again permits us to reject the hypothesis of a poor fit of the model in the population.

The remainder of Table 4 shows tests of three specific null hypotheses. The first confirms that all of the traits do indeed show significant male-female differences in means: the significant chi-square differences indicate rejection of the constraint that the means be equal. The second tests the mean differences across ages in Table 2. Again, significant chi-square increases on all four traits result from imposing equality of means across age groups. The third tests the departure of the age regressions from zero. For all four traits, chi square increases significantly when the regression is fixed to zero. The models themselves are based on covariances and unstandardized regressions, but for convenience of comparison the standardized fitted regression is shown in the Table; it can be considered an average within-group correlation over the six groups.

3.2. Longitudinal comparisons of means and age regressions

Cross-sectional comparisons such as those made in Tables 2–4 involve some degree of ambiguity — do the changes involve changing ages or changing times? The consistency of changes within and between groups offers some reassurance on this score, as the times of testing the groups were scrambled relative to their ages — the youngest group was in fact tested between the middle and the oldest. Further reassurance can be provided by examining the two longitudinal samples.

As previously noted, the intermediate-age group in Tables 2 and 3 was a follow-up of a 1980 sample. The 1980 sample had received an earlier 90-item version of the Eysenck inventory. This version contained all 12 items included in the short versions of the Extraversion and Neuroticism scales, and the 11 items scored for Lie, but only eight of the 12 Psychoticism items. Thus we can make longitudinal comparisons over roughly an 8-year interval that are analogous to the cross-sectional comparisons in Tables 2–4, for the same E, N, and L scales, and for an 8-item P scale. The P scale has a slightly different flavor in the longer than in the shorter version. The four added items (numbers 2, 14, 26, and 43 in Eysenck et al., 1985, Appendix 2) express an active preference for social isolation, for being a loner — in contrast with a simple disregard for social conventions, a theme present in both the eight- and 12-item versions of the scale.

A second longitudinal comparison was also possible, due to the fact that, as noted, a number of the older members of the intermediate sample also participated, some 6 years later, in the old-age sample; most had filled out questionnaires in the original 1980 study as well. In this case we can compare their scores on the eight-item P scale and the E, N, and L scales across a 14-year interval later in the life span. This thrice-tested group constitutes a subset from the older end of the twice-tested group described in the preceding paragraph. To avoid duplication, this subset is excluded from the twice-tested group in the tables to follow: that is, the 'younger' and 'older' groups in these tables have no overlapping members. For simplicity, the descriptive tables show data for the two sexes combined, although in the statistical modeling they were kept separate as before.

Table 5 shows the comparisons of means and standard deviations for the retested groups. First, the variabilities — the standard deviations — do not differ much from test to retest for either the earlier or later comparisons, and do not differ much between them. A possible exception is P, for which means and variances tend to be correlated. Again, we use the transformed version Pq for model fitting.

In the case of the means, for the younger group, and for the first comparison for the older group, the longitudinal differences track the cross-sectional ones reported in Tables 2 and 3: P, E, and N decrease and L increases. However, at the last age, things are different. There is still a decrease in N, but P and E do not decrease — if anything, they increase slightly — and L declines instead of continuing to increase.

Table 6 shows within-group correlations of Eysenck scales with age at the various ages for the two retested groups. These results are quite consistent with the cross-sectional results reported earlier: negative correlations with age for P, E, and N, and positive for L at all ages. The persons tested repeatedly are not different from the larger group from which they come when comparisons are made across individuals; it is only changes within individuals at the later age that are distinctive.

Table 5
Means and standard deviations on Eysenck scales for two groups of retested individuals[a]

Group and testing	Mean age	Means				SDs				No.
		P	E	N	L	P	E	N	L	
Younger										
1st	29.3	0.88	7.18	4.98	4.42	1.04	3.36	3.19	2.54	4585
2nd	37.6	0.76	6.88	4.57	5.24	0.93	3.54	3.20	2.62	4585
Older										
1st	47.5	0.66	6.53	4.75	5.46	0.84	3.50	3.28	2.57	1228
2nd	55.7	0.63	6.35	4.42	6.19	0.79	3.52	3.25	2.58	1228
3rd	61.9	0.75	6.42	3.83	5.54	0.87	3.49	3.11	2.52	1228

[a] P = Psychoticism, E = Extraversion, N = Neuroticism, L = Lie. No. = average number of individuals.

Table 6
Within-group correlations with age for Eysenck scales in two groups of retested individuals[a]

Testing	P	E	N	L	No.
Younger					
1st	−0.12	−0.13	−0.05	0.30	4585
2nd	−0.10	−0.11	−0.04	0.25	4585
Older					
1st	−0.05	−0.04	−0.08	0.21	1228
2nd	−0.10	−0.01	−0.05	0.22	1228
3rd	−0.05	−0.03	−0.05	0.27	1228

[a] P = Psychoticism, E = Extraversion, N = Neuroticism, L = Lie. No. = Average number of individuals.

Table 7 provides formal model-fitting tests of the age trends in Tables 5 and 6. The models include the same assumptions as those fitted earlier to the cross-sectional data: constant regressions with age, constant sex differences across ages, and constant residual variances. The overall fits of the models, as indicated by the RMSEAs, are reasonable. On the whole, they are good for E, N, and L, and marginal for P.

Following the overall model fit, the first chi-square difference test in the table show that the changes between the mean ages of 29 and 38 in the twice-tested sample — the average decreases in P, E, and N, and the increase in L — are all highly statistically significant. The corresponding test for ages 48–56 in the lower half of the table shows similar but weaker results (recall that the sample size is considerably less here); the decrease in P is not statistically significant. In the next test in the lower half of the table the reversals for P and L between ages 56 and 62 are shown both to be significant, as is the continued decrease in N; the change in E is not.

The last rows in both parts of the table indicate that the within-group age regressions are generally significant — only that for E in the thrice-tested group falls short of statistical dependability.

3.3. *Age trends in genetic and environmental parameters*

We turn now to a consideration of heritabilities.

Table 8 presents twin correlations for MZ and DZ twins for the three Australian Twin Registry samples of Tables 2 and 3. These are simple Pearson correlations between the scores of the first- and second-listed twins of a pair as they occur in the data files.

Table 7
Model fitting to means and covariances of age and personality scales for males and females in retested groups[a]

Test	Pq	E	N	L
Group tested twice at mean ages 29 and 38				
Fit of model				
χ^2, 9df	165.15*	39.30*	10.00	48.16*
RMSEA	0.089	0.038	0.007	0.043
UL	0.100	0.051	0.025	0.056
Do means differ from 29 to 38? χ^2_{diff}, 1df	52.17*	55.65*	105.23*	603.27*
Is age regression zero? χ^2_{diff}, 1df	70.41*	78.96*	17.00*	445.15*
Standardized age regression	−0.10	−0.12	−0.05	0.27
Group tested three times at mean ages 48, 56 and 62				
Fit of model				
χ^2, 16df	60.70*	11.30	20.46	24.16
RMSEA	0.072	0.000	0.021	0.030
UL	0.090	0.027	0.045	0.051
Do means differ from 48 to 56? χ^2_{diff}, 1df	1.09	7.26*	21.33*	158.06*
Do means differ from 56 to 62? χ^2_{diff}, 1df	21.41*	2.27	76.59*	132.36*
Is age regression zero? χ^2_{diff}, 1df	8.28*	0.96	6.16*	89.56*
Standardized age regression	−0.06	−0.03	−0.06	0.24

[a] Models allow means to differ between the sexes, but assume constant regressions with age, residual variances, and sex differences over age. Pq = square-root-transformed P. RMSEA = root mean square error of approximation. UL = upper limit of 90% confidence interval for RMSEA. *$p < 0.05$.

Table 8
Male and female twin correlations for Eysenck scales in three twin samples[a]

Group	Males					Females				
	P	E	N	L	Pairs	P	E	N	L	Pairs
Young										
MZ	0.22	0.52	0.36	0.38	255	0.40	0.42	0.37	0.44	453
DZ	0.12	0.23	0.18	0.27	164	0.20	0.15	0.11	0.28	312
OS	−0.13	0.11	0.12	0.21	344	−0.13	0.11	0.12	0.21	344
Middle										
MZ	0.37	0.51	0.38	0.28	327	0.35	0.48	0.44	0.51	685
DZ	0.24	0.24	0.02	0.34	195	0.20	0.20	0.25	0.33	420
OS	0.12	0.15	0.14	0.18	460	0.12	0.15	0.14	0.18	460
Old										
MZ	0.30	0.57	0.27	0.44	140	0.30	0.46	0.44	0.51	470
DZ	0.17	0.20	0.24	0.20	68	0.18	0.10	0.19	0.29	250
OS	0.14	0.12	0.10	0.12	224	0.14	0.12	0.10	0.12	224

[a] P = Psychoticism, E = Extraversion, N = Neuroticism, L = Lie. Pairs = average number of twin pairs. MZ = monozygotic twins, DZ = same-sex dizygotic twins, OS = opposite-sex dizygotic twins (shown twice, under both males and females).

Twin correlations are given separately by sex for MZ, same-sex DZ, and opposite-sex DZ pairs, although it will be observed that some of the sample sizes are becoming fairly small — for example, the 68 male same-sex DZ pairs in the older sample. (For convenience of comparison, the opposite-sex pairs are shown in the table under both 'males'' and 'females', although these of course are the same pairs.)

On the whole, the correlations appear roughly comparable across the two sexes and across the three age groups. Clearly the MZ correlations tend to exceed the DZ correlations, suggesting genetic influences on these traits. For Extraversion and Neuroticism, the DZ correlations are usually less than or roughly equal to one-half the corresponding MZ correlations, suggesting little influence of shared environment on these traits. A number of the DZ correlations for P and L are greater than one-half the corresponding MZ correlations, suggesting that shared environments may be more of a factor for these scales.

The opposite-sex DZ correlations appear to run slightly lower than the same-sex DZ correlations, suggesting the possible presence of sex limitation, although not to any marked degree. For simplicity in model fitting, and because our main concern in this paper is age trends in heritabilities, we will restrict our analysis to the same-sex pairs.

Table 9 presents the results of fitting a simple heredity-environment model to twin covariance matrices for the 12 groups of same-sex twins. The model is a standard biometric one, including the effects of additive genes (h^2), shared environment (c^2), and a residual (e^2) comprising environmental influences unique to the individual, developmental anomalies, gene-environment interactions, random trait fluctuations, and measurement error.

The first rows of Table 9 show a test of this model under the assumption of equality of *h, c*, and *e* across the three ages and the two sexes. For Extraversion and Neuroticism this simple model fits extremely well: in each case the chi square is less than its degrees of freedom and the RMSEA is estimated as zero, with even its upper confidence limits representing excellent fits. The h^2s are estimated as 0.47 and 0.40, respectively, and the c^2s as zero. The transformed Psychoticism scale also fits quite well, although the heritability estimate is lower, at 0.28. There is a small estimate for c^2 in this case, 0.04, although this could represent a chance departure from zero in the population ($\chi^2 = 0.73$, below). The fit for L, though somewhat less good, is still acceptable (RMSEA = 0.045); the upper limits of the confidence intervals indicate that we can reject the hypothesis of a poor fit in the population for any of the four traits. L also yields a heritability of 0.28, and its c^2 estimate of 0.16 indicates a significant influence of shared environment on this scale ($\chi^2 = 9.87$).

The last two rows of Table 9 ask whether there are significant age or sex differences in the genetic or environmental parameters. The answer is 'no' to both for P, E, and N, and 'yes' to both for L.

For L, Table 10 shows h^2, c^2, and e^2 values in the six age and sex groups. The difference between the young and the old groups is similar for both sexes: a moderate h^2 increases, and a moderate c^2 drops to zero. The middle age group is more complicated. The c^2s continue moderate at these ages, but the h^2s show a striking sex difference. The male h^2 inexplicably drops to zero, whereas the female h^2 is intermediate between those in the young and old groups. The anomalous male h^2 reflects the corresponding correlations in Table 8, where the DZ r exceeds the MZ r,

Table 9
Model fitting to covariance matrices for same-sex twin pairs in three age groups[a]

Test	Pq	E	N	L
Fit of model				
χ^2, 33df	36.16	27.32	25.72	49.21*
RMSEA	0.021	0.000	0.000	0.045
UL	0.048	0.028	0.027	0.066
h^2	0.28	0.47	0.40	0.28
c^2	0.04	0.00	0.00	0.16
e^2	0.68	0.53	0.60	0.56
Is $c^2 = 0$? χ^2_{diff}, 1df	0.73	0.00	0.00	9.87*
Sex differences? χ^2_{diff}, 3df	5.03	6.14	5.15	21.15*
Age differences? χ^2_{diff}, 6df	11.90	2.60	3.10	17.59*

[a] Pq = square-root-transformed P. RMSEA = root mean square error of approximation. UL = upper limit of 90% confidence interval for RMSEA.

Table 10
Genetic and environmental parameters for L by age and sex groups

Group	h^2	c^2	e^2
Young, male	0.21	0.17	0.62
Young, female	0.26	0.17	0.57
Middle, male	0.00	0.30	0.70
Middle, female	0.36	0.14	0.50
Old, male	0.43	0.00	0.57
Old, female	0.43	0.00	0.57

representing the only such reversal in the Table. We have no explanation for this, if it is not just an aberration of sampling. The *N*s of 195 and 327 are not huge, but they are by no means minuscule. We have scores for most of these individuals from eight years before, so we can ask whether they showed this same reversal then. The answer is that they did not. The DZ correlations were about the same on the two occasions — 0.34 and 0.33 — but the MZ correlation was 0.43 instead of the present 0.28. We conclude that it would be well to await replication before making much of this apparent sex difference.

For all four traits, then, what does this analysis say about the 'growing apart' found by McCartney et al. (1990) in their survey? This would be shown by a decline in c^2 with age. For P, E, and N, the same model fit at all three ages. There is a decline in c^2 for L in the oldest group, but this represents an age range not included in the McCartney et al. review. More likely, the growing apart that they observed was chiefly a phenomenon of the adolescent years, as twins move away from their families and each other to establish separate identities and places in the world, a process presumably stabilizing by adulthood.

For Extraversion and Neuroticism in the Finnish study, c^2s were estimated at zero at all adult ages, consistent with our result. The Finnish h^2s decreased somewhat for both traits after the youngest group (ages 18–23), remaining fairly stable thereafter through the adult years, with a slight drop after age 42 for E.

3.4. Reliability and stability

In comparing the various estimates across traits, one needs to consider the relative reliabilities with which they are measured. Table 11 provides this information, in the form of internal consistency estimates of reliability (Cronbach alphas).

As is evident from Table 11, E tends to be the most reliable, with N close behind, and the slightly shorter L scale a little lower. P is markedly less consistent than the others, even in its 12-item version, suggesting that this scale is less satisfactory from a psychometric point of view. Its lower heritability estimate in Table 9 may in part reflect this lower reliability. In general, estimates of reliability appear to be consistent across the various age samples.

The estimates in Table 11 are based on internal consistencies assessed at a given time. Table 12 presents cross-time correlations for the longitudinal samples. Because these are across intervals of 6–14 years, they jointly reflect reliability of measurement and trait stability. A comparison of Table 11 and Table 12 suggests that the stabilities must be quite high. Indeed, in the case of P, its

Table 11
Reliabilities of Eysenck scales[a]

Sample	P12	P8	E12	N12	L11	No.
Young	0.50		0.86	0.72	0.69	3108
Middle	0.47	0.30	0.87	0.83	0.72	3943
Old	0.38	0.25	0.84	0.83	0.72	2347
1980		0.40	0.85	0.81	0.73	7613

[a] P12 = 12-item P scale (etc.). Reliabilities are Cronbach alphas calculated over individuals having no missing data. No. = average number of individuals.

Table 12
Correlations across time for retested individuals on Eysenck scales[a]

Testing	P	E	N	L	No.
Younger	0.42	0.71	0.64	0.64	4585
Older					
1st to 2nd	0.44	0.80	0.71	0.70	1228
2nd to 3rd	0.44	0.80	0.74	0.71	1228
1st to 3rd	0.31	0.75	0.66	0.69	1228

[a] P = Psychoticism (eight-item scale), E = Extraversion, N = Neuroticism, L = Lie. No. = average number of individuals.

6–8 year retest correlations are higher than its internal consistency reliability, suggesting that this must be a heterogeneous scale in terms of content (Heath, Jardine, Eaves & Martin, 1988; Heath & Martin, 1990). The stabilities for E and N run a little higher than the 6-year stabilities of 0.48 to 0.74 obtained with slightly shorter E and N scales in the Finnish twin study (Viken et al., 1994), and those for P, E, N, and L are a little below the 6-year stabilities of 0.61 to 0.84 obtained in a Dutch study using longer scales (Sanderman & Ranchor, 1994).

Both the stabilities and the internal consistencies are as high for the oldest group as for the others. Thus it is clear that the anomalous changes in the older group are not simply reflecting less consistency of personality at this age.

4. Discussion and conclusions

The cross-sectional data in these twin samples, and the longitudinal data at most ages, tell a simple story. The Eysenckian traits of Psychoticism, Extraversion, and Neuroticism decline over age during adulthood, and the conformity measured by Lie increases.

The age trends for E and N are consistent with those reported by Costa et al. (1986) for Extraversion and Neuroticism in a US national sample of about 10,000 adults between 32 and 88 years of age. The correlations with age obtained in that study were −0.16 for Extraversion and −0.12 for Neuroticism.

The Finnish twin study agreed in part: Extraversion means declined with age for both sexes and Neuroticism did for women, but there was a more complex pattern for N in men, with both increases and decreases (Viken et al., 1994). A smaller Minnesota longitudinal twin study found declines between ages 20 and 30 in a measure similar to N, but failed to find significant differences in a measure similar to E (McGue, Bacon & Lykken, 1993).

The trends for all four Eysenckian traits appear to be consistent with the correlations obtained by McCrae et al. (1999) across five countries with differing cultures (Germany, Italy, Portugal, Croatia, and Korea). These authors used translated versions of the NEO-PI-R (Costa & McCrae, 1992), a questionnaire which measures the so-called Big Five personality factors. They found declines with age for three of the factors, Neuroticism, Extraversion, and Openness, and increases with age for two factors, Conscientiousness and Agreeableness. Neuroticism and Extraversion show the same trends as in the present study. Eysenck's P factor may be considered as a combination of Conscientiousness and Agreeableness, both reversed, which would make these changes consistent with the present findings. Finally, the tendency toward an increase in social conformity with age, as represented by increasing L scale scores, seems consistent with a decline in the independence inherent in Openness.

Thus the cross-sectional age trends in the present study are consistent with those that have been reported for adult non-twin populations in the US, Western and Eastern Europe, and Asia.

However, the longitudinal changes over a 6-year period for the over-50 group were markedly at odds with this simple picture. One speculation might be that in this age range, as we get beyond the ages when natural selection was effective during human prehistory, the factors governing personality change themselves change. In the cognitive domain, Finkel, Pedersen, McGue, and McClearn (1995) reported lower heritabilities and a changed factor structure among Swedish twins past age 65.

Turning to heredity-environment parameters: Individual differences in Extraversion, Neuroticism, and Psychoticism appeared in our data to be substantially influenced by the genes, and stably so over the adult years. Individual differences in conformity (Lie) showed the effect of shared environments as well as genes, and showed changes with age. For E and N, these results are supported by the Finnish twin study, with the exception of some decreases in heritability early and late in adulthood (Viken et al., 1994). The Minnesota twin study also found a decline in heritability in early adulthood for negative emotionality (N) (McGue et al., 1993).

The 'growing apart' of twins observed by McCartney et al. (1990) in their meta-analysis may have been largely a feature of late adolescence. It was not in evidence during the bulk of adulthood in the Australian twin samples. For social attitudes, Eaves et al. (1997) reported a sharp decrease in correlations for DZ twins in their early 20s, with a persistence of resemblance among MZs, and reasonable stability of the correlations thereafter through adulthood, except for a downturn in both groups in the oldest age interval (75+ years). On the other hand, in a large sample of Swedish twin pairs there was an increasing proportion of twin pairs whose scores differed markedly on a Neuroticism scale as ages increased from 16 to 48 years (see Rose, Koskenvuo, Kaprio, Sarna & Langinvainio, 1988). This was true for both MZ and DZ pairs, and for males and females, suggesting that McCartney et al.'s 'growing apart' of twins can apparently continue through a substantial part of adulthood for at least some traits in some populations.

In the present study, the tendency toward less effect of shared environment on the Lie scale in the over-50 age group is interesting. If this scale is interpreted as in part a measure of social conformity, one interpretation might be that the critical event here is the departure of the children from the home. Until that occurs, one of the responsibilities felt by most parents is to set good examples for their children by conforming to the standards acquired in their own homes of origin. After the emptying of the nest, the behavior of the parents becomes more a matter of their individual choice. Such a shift could result in lower c^2s in the over-50 group, and possibly in higher h^2s. The anomalous results for the males in the middle age group, if real, remain a puzzle.

To summarize: The Australian twin data show consistent age changes in means on the Eysenck scales, both cross-sectionally and longitudinally (except for the oldest group for the latter). With a few exceptions these trends, decreases over time in P, E, and N, and increases in L, are consistent with those observed using other instruments in other populations, and may be presumed to constitute a fairly general feature of adult personality. Genetic contributions to individual differences in E and N were substantial and stable over time during adulthood. Somewhat lower heritabilities for P may partly reflect its less satisfactory measurement. Shared environment made an appreciable contribution to L, dropping off in late adulthood, at a point where the longitudinal data on means also suggested changes in age trends.

References

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long, *Testing structural equation models* (pp. 136–162). Thousand Oaks, CA: Sage.

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

Costa, P. T. Jr, McCrae, R. R., Zonderman, A. B., Barbano, H. E., Lebowitz, B., & Larson, D. M. (1986). Cross-sectional studies of personality in a national sample: 2. Stability in neuroticism, extraversion, and openness. *Psychology and Aging*, 1, 144–149.

Duffy, D. L. (1994). Asthma and allergic diseases in Australian twins and their families. Unpublished PhD thesis, University of Queensland, Brisbane, Australia.

Eaves, L., Martin, N., Heath, A., Schieken, R., Meyer, J., Silberg, J., Neale, M., & Corey, L. (1997). Age changes in the causes of individual differences in conservatism. *Behavior Genetics*, 27, 121–124.

Eysenck, H. J., Eysenck, S. B. G., & Barrett, P. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences*, 6, 21–29.

Finkel, D., Pedersen, N. L., McGue, M., & McClearn, G. E. (1995). Heritability of cognitive abilities in adult twins: Comparison of Minnesota and Swedish data. *Behavior Genetics*, 25, 421–431.

Heath, A. C., Jardine, R., Eaves, L. J., & Martin, N. G. (1988). The genetic structure of personality: I. Phenotypic factor structure of the EPQ in an Australian sample. *Personality and Individual Differences*, 9, 59–67.

Heath, A. C., & Martin, N. G. (1990). Psychoticism as a dimension of personality: A multivariate genetic test of Eysenck and Eysenck's Psychoticism construct. *Journal of Personality and Social Psychology*, 58, 111–121.

Lykken, D. T., Tellegen, A., & DeRubeis, R. (1978). Volunteer bias in twin research: The rule of two-thirds. *Social Biology*, 25, 1–9.

Macaskill, G. T., Hopper, J. L., White, V., & Hill, D. J. (1994). Genetic and environmental variation in Eysenck Personality Questionnaire scales measured on Australian adolescent twins. *Behavior Genetics*, 24, 481–491.

Martin, N., & Jardine, R. (1986). Eysenck's contributions to behaviour genetics. In S. Modgil, & C. Modgil, *Hans Eysenck: Consensus and controversy* (pp. 13–47). Philadelphia: Falmer Press.

McCartney, K., Harris, M. J., & Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychological Bulletin*, 107, 226–237.

McCrae, R. R., Costa, P. T. Jr, de Lima, M. P., Simões, A., Ostendorf, F., Angleitner, A., Marusic, I., Bratko, D., Caprara, G. V., Barbaranelli, C., Chae, J.-H., & Piedmont, R. L. (1999). Age differences in personality across the adult life span: Parallels in five cultures. *Developmental Psychology*, 35, 466–477.

McGue, M., Bacon, S., & Lykken, D. T. (1993). Personality stability and change in early adulthood: A behavioral genetic analysis. *Developmental Psychology*, 29, 96–109.

Rose, R. J., Koskenvuo, M., Kaprio, J., Sarna, S., & Langinvainio, H. (1988). Shared genes, shared experiences, and similarity of personality: Data from 14,288 adult Finnish co-twins. *Journal of Personality and Social Psychology*, 54, 161–171.

Sanderman, R., & Ranchor, A. V. (1994). Stability of personality traits and psychological distress over six years. *Perceptual and Motor Skills*, 78, 89–90.

Steiger, J. H. (1989). *EzPATH: Causal modeling*. Evanston, IL: Systat Inc.

Viken, R. J., Rose, R. J., Kaprio, J., & Koskenvuo, M. (1994). A developmental genetic analysis of adult personality: Extraversion and neuroticism from 18 to 59 years of age. *Journal of Personality and Social Psychology*, 66, 722–730.