Journal of Personality and Social Psychology
2010, Vol. 98, No. 6, 995–1008

0022-3514/10/$12.00 DOI: 10.1037/a0019558

Sources of Cumulative Continuity in Personality: A Longitudinal Multiple-Rater Twin Study

Christian Kandler, Wiebke Bleidorn, and Rainer Riemann
Bielefeld University

Frank M. Spinath
Saarland University

Wolfgang Thiel and Alois Angleitner
Bielefeld University

This study analyzed the etiology of rank-order stability and change in personality over a time period of 13 years in order to explain cumulative continuity with age. NEO five-factor inventory self- and peer report data from 696 monozygotic and 387 dizygotic twin pairs reared together were analyzed using a combination of multiple-rater twin, latent state-trait, and autoregressive simplex models. Correcting for measurement error, this model disentangled genetic and environmental effects on long- and short-term convergent valid stability, on occasional influences, and on self- and peer report-specific stability. Genetic factors represented the main sources that contributed to phenotypic long-term stability of personality in young and middle adulthood, whereas change was predominantly attributable to environmental factors. Phenotypic continuity increased as a function of cumulative environmental effects, which became manifest in stable trait variance and decreasing occasion-specific effects with age. This study's findings suggest a complex interplay between genetic and environmental factors resulting in the typical patterns of continuity in personality across young and middle adulthood.

Keywords: personality, continuity, genetics, environment, twin study

The meta-analysis by Roberts and DelVecchio (2000) profoundly established that rank-order stability in personality increases across the life course until it reaches its peak in later adulthood after age 50 (Caspi & Roberts, 2001; Fraley & Roberts, 2005). This increasing continuity in personality proved to be a robust finding across self- and other reports, independent of gender and the specific trait considered (e.g., Costa & McCrae, 1988; Terracciano, Costa, & McCrae, 2006). Three prominent theories provide rather conflicting etiological explanations for the cumulative rank-order continuity with age: the genetic set-point model (Carey, 2002), the genetic maturation hypothesis (McCrae et al., 2000), and a model proposing transactions between genetic and environmental factors (Caspi, Roberts, & Shiner, 2005). We analyzed personality assessments of twins on global personality traits of the five-factor model (McCrae & John, 1992) across three waves of measurement over a period of 13 years in order to test the adequacy of predictions from each of these etiological theories. Because this is the first longitudinal twin study that included self- and peer reports, we were able to generalize our findings across multiple raters demonstrating convergent validity (Campbell & Fiske, 1959).

Christian Kandler, Wiebke Bleidorn, Rainer Riemann, Wolfgang Thiel, and Alois Angleitner, Department of Psychology, Bielefeld University, Bielefeld, Germany; Frank M. Spinath, Department of Psychology, Saarland University, Saarbrücken, Germany.

Correspondence concerning this article should be addressed to Christian Kandler, Department of Psychology, Bielefeld University, Universitätsstr, 25, D-33615 Bielefeld, Germany. E-mail: christian.kandler@uni-bielefeld.de

The Genetic Set-Point Hypothesis

A number of behavioral genetic studies have led to the conclusion that genetic factors primarily contribute to stability in personality traits (e.g., McGue, Bacon, & Lykken, 1993; Viken, Rose, Kaprio, & Koskenvuo, 1994). Carey (2002) interpreted this finding in terms of a set-point model, in which environmental fluctuations are assumed to affect short-term changes (a few days, weeks, or even several months) in personality, whereas genetic factors determine individual set-points to which individuals will regress on a long-term basis. These assumptions are in accordance with the intriguing finding that parental environments seem to have no long-term influence on personality in adulthood (e.g., Kandler, Riemann, & Kämpfe, 2009; Krueger, Markon, & Bouchard, 2003).

According to this theory, variance due to individual genetic set-points does not change across time. The cumulative stability of personality across the life span should thus result exclusively from decreasing effects of environmental fluctuations with age. To the degree that more of the important life transitions occur in young adulthood (e.g., vocational training, finding a job, starting a family), this seems to be a plausible assumption. However, there is powerful evidence that personality stability decreases as the time interval between measurement occasions increases (Fraley & Roberts, 2005), reflecting long-term rank-order change that is not compatible with the genetic set-point hypothesis. If phenotypic scores get closer and closer to the genetic set point as a function of decreasing short-term environmental influences, then the correlation among scores of initial intervals of time will necessarily increase across a series of later intervals, even though the retest interval is also increasing. This prediction is not in line with the results and continuity functions presented by Fraley and Roberts.

Despite these conflicting findings, the genetic set-point model is still a quite appealing and parsimonious model of personality development. Even though the entire model seemed to be too restrictive, the specific assumption of decreasing effects of situational fluctuations might be a relevant mechanism of increasing stability and thus for personality development, which is worthwhile to study in more detail.

The Genetic Maturation Hypothesis

The five-factor theory (FFT) provides an alternative explanation of personality development proposing that both rank-order continuity and change in traits, considered as *basic tendencies* (abstract psychological potentials), are exclusively mediated by genetic factors (McCrae et al., 2000). According to this hypothesis, significant environmental effects on traits should merely result from short-term contextual influences and systematic as well as random measurement error.

Researchers of a number of biometric studies have provided some support for this genetic maturation hypothesis because they found stability to be primarily influenced by genetic factors and have even obtained evidence for the appearance of new genetic factors during young adulthood (e.g., Bratko & Butkovic, 2006; Viken et al., 1994). Actually, when measurement error is controlled for, genetic factors seem to contribute largely to individual stability and growth in Emotional Stability, Agreeableness, and Conscientiousness over a time period of 10 years in middle adulthood (Bleidorn, Kandler, Riemann, Angleitner, & Spinath, 2009). However, most findings have contradicted the hypothesis of an exclusive genetic maturation of personality because personality continuity has been found to be also attributable to environmental factors (e.g., Blonigen, Carlson, Hicks, Krueger, & Iacono, 2008; Johnson, McGue, & Krueger, 2005). Differential individual growth in Extraversion and Openness were largely due to environmental factors (Bleidorn et al., 2009).

At this point, it should be noted that previous longitudinal behavioral genetic studies on personality development have exclusively relied on self-reports. As a consequence, these studies could not provide a critical test of the genetic maturation hypothesis, because it was not possible to control for nonrandom bias (method) and random (measurement error) effects, whereas a multimethod longitudinal behavioral genetic study would address this issue. In a cross-sectional study, Riemann, Angleitner, and Strelau (1997) found that the employment of self- and peer reports lead to higher estimates of heritability in personality traits by subtraction of error and method variance. Recently, this finding could be replicated and extended on personality facets (Kandler, Riemann, Spinath, & Angleitner, in press). However, these estimates were still different from unity. Therefore, McCrae and colleagues (2000) acknowledged that the small remaining variance might include true environmental influences, including biological sources such as prenatal infections or different metabolisms, which could distinguish the development of genetically identical individuals. If very early environmental effects contribute to stable differences in personality, they should not change across the life course. On the basis of the assumptions of this *weaker* genetic maturation hypothesis, cumulative phenotypic continuity should result from cumulative genetic continuity, whereas environmental effects contribute to stability, situational fluctuations, and systematic as well as random error in personality measures.

The Gene–Environment Transaction Hypothesis

The two theories described above provide elegant and parsimonious explanations for the increasing rank-order stability of personality over the life span. However, the complete picture of findings appears to be incompatible with both the genetic set point and the genetic maturation hypothesis. First, large sample longitudinal twin studies (Pederson & Reynolds, 1998, 2002; Viken et al., 1994) have found that phenotypic variance increases with age as a function of increasing nonshared environmental effects. Furthermore, there is evidence for phenotypic stability to increase as a result of increasing environmental stability (Viken et al., 1994). In view of these findings, McCrae and Costa (2008) revised important tenets of the FFT. They postulated that personality development is determined by *biological* maturation. That is, genetic factors still play a crucial role, but the environment can also affect personality traits through biological bases (such as drugs, disease, etc.), and can thus affect personality change.

Caspi et al. (2005) proposed an integrative theory of personality development. They postulated that continuity and change result from transactions between genetic and environmental factors contributing to estimates of both genetic and environmental effects on phenotypic stability and change. The increasing continuity with age is considered as a process of developing and maintaining an identity (Roberts & Caspi, 2003). From this perspective, personality development may be best explained as a result of two mutually supportive life-course dynamics (Caspi et al., 2005): First, people select environments that are correlated with their personality traits (*social selection*); second, experiences in these contexts affect personality functioning (*social influence*), resulting in cumulating effects over the life course. On the basis of the gene–environment transaction hypothesis, phenotypic rank-order stability should increase with age as a result of an accumulation of environmental influences on trait variance leading to an increase of environmental continuity.

Aims of the Present Study

Only a handful of behavioral genetic studies of personality have estimated continuity and change longitudinally at more than two waves of measurement (e.g., Bleidorn et al., 2009; Pederson & Reynolds, 1998, 2002). Another restriction of previous behavioral genetic studies concerns the sole reliance on self-reports. In the present longitudinal study, we analyzed personality scales of the five-factor model (FFM; McCrae & John, 1992) assessed by self- and peer raters, spanning a time period of 13 years with three waves of assessment. As a consequence, we were able to answer questions about short- and long-term personality stability and change, which was necessary to test the adequacy of the predictions derived from the three conceptions introduced above.

In order to provide a critical test of the three conflicting hypotheses, we combined a multiple-informant twin model (Riemann et al., 1997; Riemann & Kandler, in press), a latent state-trait model (Steyer, Schmitt, & Eid, 1999), and a genetic simplex model (Boomsma & Molenaar, 1987). The availability of self- and peer reports of twins' personality allowed us to decompose convergent

valid (i.e., shared by self- and peer reports) variance into (a) long-term stable genetic and environmental sources (about 13 years), (b) "short-term" stable sources (about 6.5 years), and (c) occasion-specific genetic and environmental variance reflecting sources of change and situational fluctuations. We used structural equation modeling to test the three conflicting hypothesis that were tested against each other. The most complex model that allowed for genetic and environmental stability and change would reflect the gene–environment transaction hypothesis. The absence of environmental stability and change would argue for the genetic maturation hypothesis, whereas additional stability due to environmental factors would speak for the weaker position of the genetic maturation hypothesis. Finally, the most restricted model that only allowed for genetic stability (over a period of 13 years) and decreasing short-term environmental effects (<6.5 years) would provide evidence for the genetic set-point hypothesis. For testing parameter equivalence between different age groups, the complete sample was subdivided into two age subsamples representing young and middle adulthood. Because of the cumulative principle of continuity, we hypothesized stability to be larger in the older subsample.[1]

Method

Participants and procedure. Data from the first, third, and fifth wave (in the following referred to as Time 1, Time 2, and Time 3) of the Bielefeld Longitudinal Study of Adult Twins (BiLSAT; Bleidorn et al., 2009; Spinath, Wolf, Angleitner, Borkenau, & Riemann, 2005) were used. At these approximately equidistant measurement occasions, both self- and peer reports of personality were gathered between 1993 and 2008. The intervals between Times 1 and 2 averaged 6.35 years ($SD = 1.22$); between Times 2 and 3, the mean interval was 6.30 years ($SD = 0.47$). The complete time interval spanned almost 13 years.

Participants were excluded from analyses if they were younger than 16 at Time 1, because below this age, problems in understanding some items of the personality measure were reported (Borkenau & Ostendorf, 1993). Because of mortality that led to a higher degree of dropout, participants older than 75 at Time 3 (older than 62 years at Time 1) were also excluded. The resulting sample consisted of 696 monozygotic (MZ; 154 male and 542 female) and 387 dizygotic (DZ; 60 male, 213 female, and 114 opposite-sex) twin pairs at Time 1 who provided self-reports of personality. The number of participating twin pairs at each measurement occasion including dropout rates is displayed in Table 1.

Missing values were not completely at random for age using the Missing-Completely-At-Random-test (MCAR-test), $\chi^2(5, N = 1{,}083) = 37.65$, $p = .00$ (Little, 1988). *T* tests showed that dropout was larger for younger people (see also Table 1). As a consequence, a median split of age regarding available data at all points of time was used to subdivide the complete sample into two equally large age groups across measurement occasions: young (16–29 at Time 1) and middle adulthood (30–62 at Time 1). For 99.2% of the participants at the first, 98.2% at the second, and 96.4% at the last assessment, at least one peer report was available, received from peers who knew one twin but (preferably) not the cotwin very well.

Measure. The self- and peer report versions of the German Neuroticism-Extraversion-Openness-Five-Factor-Inventory (NEO-FFI; Borkenau & Ostendorf, 1993; Costa & McCrae, 1989) were administered. The NEO-FFI is a 60-item inventory designed for measuring personality on five dimensions: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. At the second measurement occasion, the NEO-FFI scales were computed from the NEO-Personality Inventory-Revised (NEO-PI-R; Costa & McCrae, 1992; Ostendorf & Angleitner, 2004), in which the NEO-FFI items are included. Cronbach's alpha for the five scales are presented in Table 2. Differences in internal consistencies between subsamples of twins and cotwins (regarding the dependence of twin siblings in a combined sample) and between age groups were not significant. The reliabilities for self-reported Agreeableness were slightly lower than for the corresponding peer reports. Openness derived from the NEO-PI-R (at Time 2) yielded somewhat higher internal consistencies compared with the NEO-FFI assessments.

Correlations between peers ranged between .38 (for Agreeableness at Time 3) and .54 (for Neuroticism at Time 3) with an average of .44. No noticeable differences were found in the degree of agreement among assessment waves, although peer raters were not necessarily the same across measurement occasions. As averaging peer reports reduces measurement error and rater bias (Hofstee, 1994), averaged peer-reported scale scores were used in all subsequent analyses. The correlations between self-reports and mean peer reports ranged between .43 (for Agreeableness at Time 3) and .62 (for Extraversion at Time 1) with an average of .53. There were no significant differences among points of time or between age subsamples. Consistently, the lowest degree of self-peer agreement was found for Agreeableness, the largest for Extraversion.

Analyses. The existence of age and gender differences can increase variance biasing twin covariance. Thus, self- and averaged peer reports were adjusted for linear and quadratic age effects as well as gender differences within each measurement occasion and age subsample using a regression procedure. This correction did not affect the age differences across measurement occasions but adjusted for age effects at a given point in time. Therefore, each measurement occasion represented the respective mean age of subsamples at this given point in time (see Table 1).

Phenotypic differential stability in self- and averaged peer reports was examined via Pearson's product–moment correlation on the basis of complete data using a pairwise deletion procedure for handling missing values. Stability coefficients were estimated for the young- and the middle-adult subsamples as well as for the short-term (between Times 1 and 2 and between Times 2 and 3) and full-term intervals (13 years) because the expectation was to

[1] At this point, it should be noted that this study was not aimed to test the three theories in their entireties, but was aimed to compare and test specific predictions regarding rank-order stability and change. It should also be noted that environmental factors may contain nongenetic psychological, sociological, biological, and historical factors. That is, this study was aimed to test the genetic maturation hypotheses with regard to rank-order stability, but it was not able to disentangle the *biological* maturation hypothesis (McCrae & Costa, 2008) versus the hypothesis of gene–environment transactions (Caspi et al., 2005) as a primary basis of personality development. It is in line with both hypotheses that predict environmental effects to accumulate across the life course, either more directly (Caspi et al., 2005) or mediated by biological processes (McCrae & Costa, 2008).

Table 1
Zygosity × Age Group Subsamples With Valid Values at Each Measurement Occasion

			Age	*N* of pairs		Dropout in %	
Age group	Time	Age: *Mdn*	*M* (*SD*)	MZ	DZ	MZ	DZ
Young adulthood	1	23	22.7 (3.9)	382	205	—	—
	2	29	28.8 (4.1)	156	95	59	54
	3	35	35.2 (4.3)	84	44	78	79
Middle adulthood	1	39	41.2 (9.1)	314	182	—	—
	2	46.5	48.1 (8.8)	140	88	55	52
	3	54	55.0 (8.8)	103	66	67	64
Total	1	28	31.2 (11.5)	696	387	—	—
	2	35	38.0 (11.8)	296	183	57	53
	3	44	46.2 (12.2)	187	110	73	72

Note. The dropout rates refer to Time 1, which are indicated by dashes. MZ = monozygotic twins; DZ = dizygotic twins.

find higher stabilities in the older subsample and within short-term intervals (Fraley & Roberts, 2005; Roberts & DelVecchio, 2000).

As we wanted to analyze all available data in biometric analyses, we tested whether the missing values were completely at random using the MCAR-test (Little, 1988) for each personality variable and each of the four twin data sets (young MZ and DZ as well as middle-aged MZ and DZ). MCAR-tests were not significant ($p > .05$). Thus, dropout was completely at random with reasonable certainty. Randomization of missing values was the precondition to receive unbiased results due to missing values and to analyze all available data via raw maximum likelihood modeling to detect genetic and environmental influences (Derks, Dolan, & Boomsma, 2006). This procedure is implemented in the statistical software package Mx (Neale, Boker, Xie, & Maes, 2003) and used for all biometric analyses.

By combining a multiple-informant twin model (Riemann et al., 1997), a latent state-trait model (Steyer et al., 1999), and a genetic simplex model (Boomsma & Molenaar, 1987), we were able to rule out some drawbacks that would arise with the single use of each of these models apart. For example, the latent state-trait model can be used to examine the accumulation of trait stability but is static and cannot assess the decrease in stability that arises when the time interval between measurement occasions increases. In contrast, the autoregressive simplex model can be used for the latter analysis, but it is not suitable to determine a single stable trait component that does not change with time. The combined model takes into consideration that rank-order stability increases with age and decreases with longer intervals. Furthermore, the extension of that model for twins reared together offers the possibility to disentangle genetic and environmental effects on several latent variables (see Hewitt, Eaves, Neale, & Meyer, 1988, for a description of such models). In the following, the model is described in terms of the usual notation for structural equation modeling.

On the phenotypic level, the model (see Figure 1) allowed us to decompose self- (*S*) and mean peer reports (*P*) at each point of time (rectangles in the figure) into a valid true score (τ), a method (ν), and a residual component (ε). The true score parameters can be considered valid to the degree to which self- and averaged peer raters' assessments converge. In other words, common variance in self- and peer reports reflects convergent validity (Campbell & Fiske, 1959). That is, true score parameters were corrected for self- and peer report-specific factors and random error.

On the convergent valid structure level (see the right side of Figure 1), each latent true score variable (τ) was further decomposed into a stable trait (ξ), in the following termed as *set point*, and an occasion-specific residual component (ζ). To disentangle short-term stability or—as the other side of the same coin—long-term rank-order change from the set-point component (ξ), regressions (λ_{21} and λ_{32}) were included between neighboring true score variables. In other words, the set-point factor (ξ) was modeled to explain variance due to long-term rank-order stability. The regressions (λ_{21} and λ_{32}) were modeled to consider the fact that rank-order stability may increase with age and decrease as the time interval increases, reflecting long-term rank-order change. That is, covariance between true score 1 (τ_1) and true score 2 (τ_2) may be smaller than the covariance between true score 2 (τ_2) and true

Table 2
Internal Consistency (Cronbach's α): Self- and Peer Reports at Each Measurement Occasion

	Time 1			Time 2			Time 3		
Scale	Self	Peer 1	Peer 2	Self	Peer 1	Peer 2	Self	Peer 1	Peer 2
Neuroticism	.85	.85	.85	.87	.87	.87	.88	.88	.86
Extraversion	.80	.80	.80	.82	.81	.79	.82	.79	.79
Openness	.63	.64	.62	.71	.70	.69	.61	.61	.60
Agreeableness	.69	.78	.78	.69	.77	.79	.71	.79	.80
Conscientiousness	.82	.84	.85	.82	.86	.85	.79	.85	.86

Note. Statistics are based on the complete sample ($N_{\text{time 1}} = 2{,}086$; $N_{\text{time 2}} = 796$; $N_{\text{time 3}} = 564$).

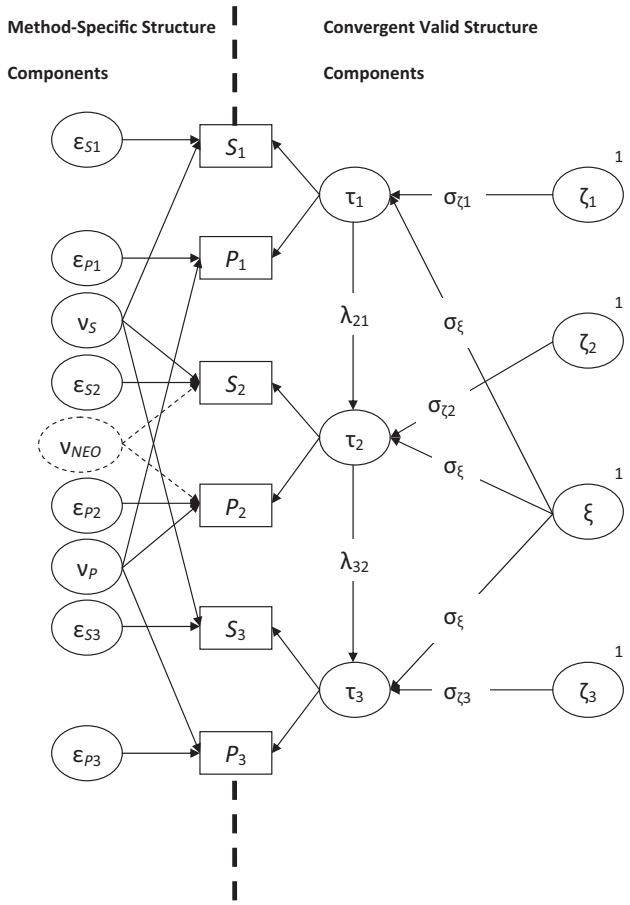

Figure 1. Full phenotypic structural equation model. S = self-report; P = peer report; indices 1, 2, and 3 = points of time; τ = convergent valid true score variable; ξ = set-point variable; ζ = true score residual; λ = linear regression of the true score variable on the previous true score variable; ν_S = stable self-report specific factors; ν_P = stable peer report specific factors; ε = measurement error; ν_{NEO} = method error with regard to the fact that NEO-FFI scales were computed from the NEO-PI-R at time 2.

score 3 (τ_3). However, covariance between true score 1 and true score 3 may be even smaller than the covariance between neighboring true scores, because the product $\lambda_{21} \times \lambda_{32}$ ($0 \leq \lambda \leq 1$) is always smaller than λ_{21} and λ_{32}. The true score residual component (ζ) was modeled to explain variance specific to each measurement occasion.

The modeling of self- and mean peer reports allowed us to estimate method factors. In our application, these factors reflected method-specific stability (ν_S and ν_P; on the left side of Figure 1). Nonstable factors specific to self- and peer reports were confounded with measurement error (ε) in our model. In addition, a systematic method factor was modeled, accounting for differences between the administered instruments, acknowledging the fact that NEO-FFI scales were computed from the NEO-PI-R at the second measurement occasion (ν_{NEO}).

For identification of this *phenotypic model*, it is possible to fix *second-level exogenous* latent variable variances (variances in ξ, ζ_1, ζ_2, and ζ_3; marked with "1" on these latent variables in Figure 1) and fix paths (marked with unlabeled arrows in Figure 1) from each *first-level exogenous* (ν_S, ν_P, ε_{S1}, ε_{S2}, ε_{S3}, ε_{P1}, ε_{P2}, and ε_{P3}) and *endogenous* (τ_1, τ_2, and τ_3) latent variables in order to estimate variance components of all *exogenous* variables (ξ, ζ_1, ζ_2, ζ_3, ν_S, ν_P, ε_{S1}, ε_{S2}, ε_{S3}, ε_{P1}, ε_{P2}, and ε_{P3}) and regressions (λ_{21} and λ_{32}). The squares of standardized loadings reflect the respective latent variable variances (σ^2_ξ, $\sigma^2_{\zeta 1}$, $\sigma^2_{\zeta 2}$, and $\sigma^2_{\zeta 3}$). However, it is also possible to fix second-level paths (aside from regressions) in order to estimate second-level variable variances. The results should be the same. The ν_{NEO} factor, which reflects nongenetic instrument-specific biases, is not identified in the *phenotypic model*. However, it is identified in a twin model or a multigroup model by equalizing this parameter across twins and (or) groups.

The next step was the extension of the *phenotypic model* to a *biometric model* (see Figure 2). Genetic (G) and environmental (E) effects were disentangled on every latent variable and regressions, except on random error (ε) and the inventory method factor (ν_{NEO}). Random error is defined to be neither consistent over different methods and occasions of measurements nor correlated within twin pairs. Effects of the inventories should be equal across MZ and DZ twins, self- and peer raters, young and middle adulthood. Thus, both components cannot be affected by genetic factors. However, the amount of genetic and environmental effects on self- and peer report method factors were estimated, which might reflect stable substantial rater-specific components of personality (e.g., self-concept, social consequences of behavior) or (and) artificial rater biases (e.g., leniency, self-enhancement).

In our application, all *exogenous* latent variable (G and E) variances were fixed to one in order to obtain estimates for all paths (σ_G and σ_E) *on the biometric level* (marked with dotted lined arrows in Figure 2). Biometric variance components were computed from the squares of these freely estimated path coefficients (σ^2_G and σ^2_E). Phenotypic components (σ^2_P) were computed from the sum of corresponding biometric components (see Neale & Maes, 2004):

$$\sigma^2_P = \sigma^2_G + \sigma^2_E. \quad (1)$$

As already mentioned, such decompositions were conducted with all phenotypic parameters (e.g., $\lambda_P = \lambda_G + \lambda_E$) except with random error and the inventory method factor. According to quantitative genetic theory, genetically identical (MZ) twins share 100% and fraternal (DZ) twins share on average 50% of their segregating genes. Cross-twin (cross i and j, see Figure 2) covariance for MZ twins (σ_{MZ}, genetic correlation: $\gamma = 1$) is equivalent to the genetic variance:

$$\sigma_{MZ} = \sigma^2_G. \quad (2)$$

And cross-twin covariance for fraternal (DZ) twins (σ_{DZ}, genetic correlation: $\gamma = .50$) is equivalent to a half of the genetic variance:

$$\sigma_{DZ} = [.5]\sigma^2_G. \quad (3)$$

From this it follows that environmental effects (σ^2_E) are implicated to the degree that MZ twins differ from one another. This genetically informative model may further be extended to nonadditive genetic influences or environmental effects shared by twin siblings (see Pederson & Reynolds, 1998, 2002, for a description of biometric common factor/simplex models). Because of inconsistent findings about nonadditive genetic effects across different studies and different methods of assessment, and considering the lack of power to detect

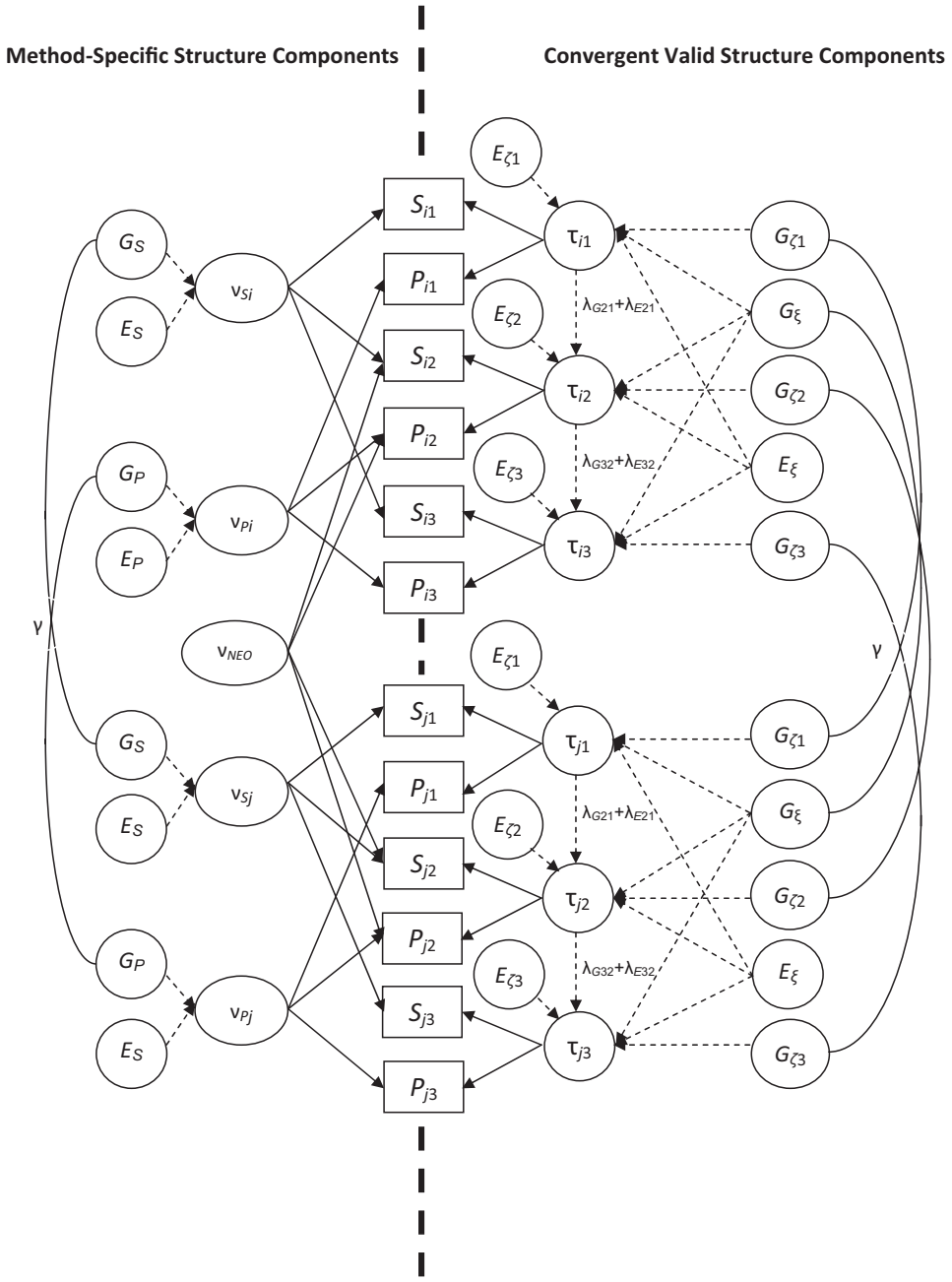

Figure 2. Full biometric structural equation model. G = genetic factors; E = environmental factors; S = self-report; P = peer report; indices i and j = twin and cotwin; indices 1, 2, and 3 = points of time; τ = convergent valid true score variable; ξ = set-point variable; ζ = true score residual; λ = linear regression of the true score variable on the previous true score variable; ν_S = stable self-report specific factors; ν_P = stable peer report specific factors; ν_{NEO} = method error with regard to the fact that NEO-FFI scales were computed from the NEO-PI-R at time 2; γ = 1.0 for MZ twins and 0.5 for DZ twins; for a better readability labels of path coefficients and measurement residual variables reflecting random error are not shown.

nonadditive genetic effects in the classical twin design (Kandler et al., 2009; Keller, Coventry, Heath, & Martin, 2005; Riemann et al., 1997), genetic effects were assumed to be additive. Environmental effects reflected sibling-specific influences of experiences referred to as nonshared environment, because environmental effects on personality shared by siblings can be assumed to be negligible (Bouchard & Loehlin, 2001; Yamagata et al., 2006).

The complete structure model on the right side of Figure 2 reflects the gene–environment transaction model. More restricted models that reflect the two other conceptions (see Figures 3a and

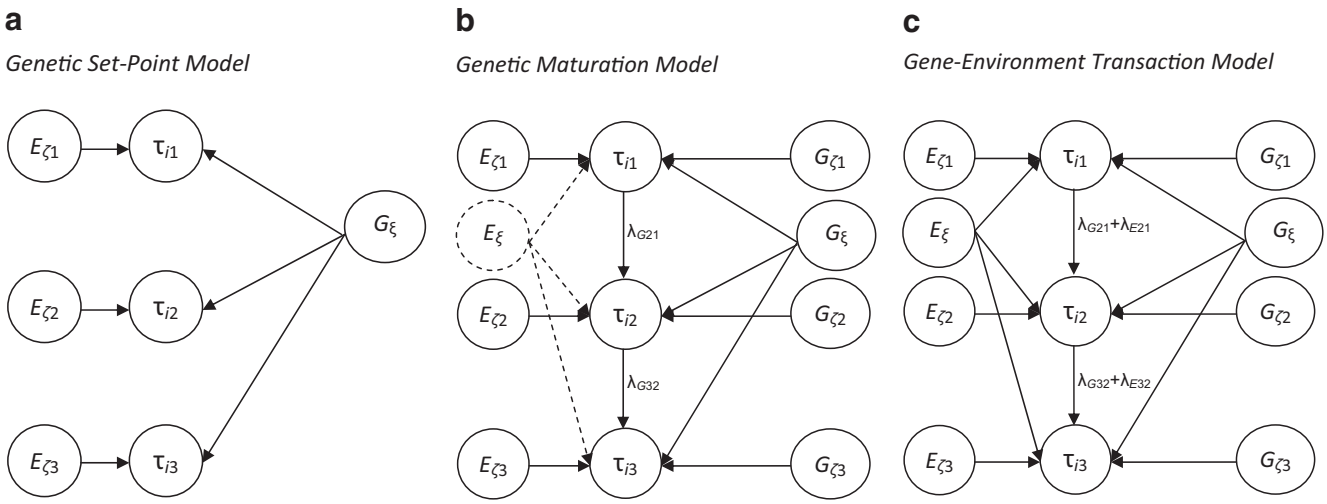

Figure 3. Conceptual structure models. Note, these models only represent portions of the full model presented in Figure 2. a: the genetic set-point model. b: the genetic maturation model. c: the gene–environment transaction model. G = genetic factors; E = environmental factors; indices 1, 2, and 3 = time 1, 2, and 3; τ = convergent valid true score variable; ξ = set-point variable; ζ = true score residual; λ = linear regression of the true score variable on the previous true score variable; further description in the text.

3b) are nested in that model, and in turn the genetic set-point model (see Figure 3a) is nested in all other models (see Figures 3b and 3c). Thus, the models were compared via −2 log-likelihood (−2*LL*) ratio tests (Neale et al., 2003). Furthermore, model modifications were tested that reflect alteration of the three conceptual models. For example, we could differentiate between a strong and a weak position of the genetic maturation hypothesis. Compared with the strong position model, the weak model allows for environmental or nongenetic biological set-point effects (McCrae et al., 2000; marked with dotted lines in Figure 3b).

Prior to model comparisons, the significance of the inventory method factor was tested. Comparing models, we began with the most parsimonious model, the genetic set-point model (see Figure 3a: $E_\xi = G_{\zeta 1} = G_{\zeta 2} = G_{\zeta 3} = \lambda_{G21} = \lambda_{E21} = \lambda_{G32} = \lambda_{E32} = 0$), as the baseline model and compared it with more complex models. First, the baseline model was compared with a model allowing for a genetic set point in the presence of genetic change (see Figure 3b: $E_\xi = \lambda_{E21} = \lambda_{E32} = 0$). This model reflects the strong position of the genetic maturation hypothesis. Then, the significance of an additional environmental set-point variable was tested ($\lambda_{G21} = \lambda_{E21} = \lambda_{G32} = \lambda_{E32} = 0$) as well as environmental change ($E_\xi = \lambda_{G21} = \lambda_{G32} = 0$) in presence of a genetic set point. Proceeding with this bottom-up strategy, more complex models were compared with nested models. For example, the strong genetic maturation model could be compared with the weak genetic maturation model that allowed for an environmental set point (see Figure 3b: $\lambda_{E21} = \lambda_{E32} = 0$). All reduced models were nested in the full model reflecting the gene–environment transaction model (see Figure 3c).

A four-group twin model (young and middle-aged MZ and DZ twins) was used to examine differences in convergent valid parameter estimates between the two age subsamples. For all model comparisons, self- and peer report-specific as well as random effects were freely estimated. After the best fitted structural equation model was identified, the equivalence of set-point variance components was tested for between the young and the middle-aged twins. Finally, phenotypic, genetic, and environmental continuity coefficients were computed, which were corrected for measurement error and method-specific effects. These coefficients were computed for each age group and for the short-term (between true scores 1 and 2 as well as between true scores 2 and 3) and long-term intervals (between true scores 1 and 3).

Results

Uncorrected phenotypic rank-order stability coefficients for self-reports and mean peer reports show apparent differences between the young and middle-aged subsamples (see Table 3). Across all personality variables, stability increased with time and age and decreased as the time interval increased. Stability coefficients were consistently smaller for mean peer reports.[2] This refers to factors affecting stability, but self-report specifically. We did not find differences between twin sibling *i* and *j* subsamples.

Multiple-group structural equation modeling is summarized in Table 4. Starting with the genetic set-point model (baseline model 0: $G_\xi + E_\zeta$), model fitting analyses did not reveal significant effects of genetic change (Model 1: $G_\xi + E_\zeta + G_\zeta + \lambda_G$) in the complete sample, but significant genetic change was found for Agreeableness and Conscientiousness in the younger subsample. That is, genetic maturation affecting rank-order change was only significant for Agreeableness and Conscientiousness in young adulthood. The inclusion of an environmental set-point variable (Model 2: $G_\xi + E_\zeta + E_\xi$) led to an increase in fit over the baseline

[2] Stability coefficients are presumably lower for averaged peer reports, because targets could have been rated by different raters at the three measurement occasions. This specificity might also account for the difference in stability of method effects between self-reports and peer ratings.

Table 3
Phenotypic Rank-Order Stabilities of Self-Reports and Averaged Peer Reports

	Time 1–2 (6.35 years)		Time 2–3 (6.30 years)		Time 1–3 (12.61 years)	
Scale	Self	Peer	Self	Peer	Self	Peer
	Young adulthood					
N	414	413	184	184	256	256
Neuroticism	.65	.44	.67	.59	.60	.51
Extraversion	.70	.55	.81	.63	.65	.50
Openness	.65	.54	.72	.56	.55	.42
Agreeableness	.60	.48	.66	.49	.47	.34
Conscientiousness	.69	.54	.64	.58	.53	.41
M	.66	.51	.70	.57	.56	.44
	Middle adulthood					
N	380	377	237	229	322	316
Neuroticism	.62	.60	.75	.69	.62	.55
Extraversion	.75	.63	.77	.64	.73	.51
Openness	.67	.61	.66	.62	.62	.55
Agreeableness	.66	.51	.73	.60	.59	.47
Conscientiousness	.67	.54	.74	.57	.63	.53
M	.68	.58	.73	.63	.64	.52

Note. Statistics are based on pairwise deletion. All correlations were significant ($p < .05$).

model ($G_\xi + E_\zeta$) across all personality variables and subsamples, except for Agreeableness in the young subsample. The model allowing for environmental change in addition to a genetic set point (Model 3: $G_\xi + E_\zeta + \lambda_E$) fitted the data significantly better than the baseline model ($G_\xi + E_\zeta$) across all personality variables and subsamples. The model allowing for both environmental set point and environmental change in the presence of a genetic set point (Model 4: $G_\xi + E_\zeta + E_\xi + \lambda_E$) improved fit over that model allowing only for an additional environmental set point (Model 2: $G_\xi + E_\zeta + E_\xi$), consistently for all personality variables. However, it did not lead to an improvement in fit over that model allowing only for environmental change (Model 3: $G_\xi + E_\zeta + \lambda_E$), except for Neuroticism, Extraversion, and Agreeableness in the older subsample. In summary, model fitting analyses provided no direct support for a genetic set-point model, because short- as well as long-term (convergent valid) environmental change was significant for all personality traits and age groups. Depending on age (young adults vs. middle adulthood) and the specific trait, different models that reflected compromises of the genetic maturation and the gene–environment transaction model were most suitable to describe the data. That is, a model that allows for a genetic set point, genetic maturation (for Agreeableness and Conscientiousness), and environmental change ($G_\xi + G_\zeta + \lambda_G + E_\zeta + \lambda_E$) provided the best fit for the data of the young adulthood subsample, whereas a model allowing for a genetic set point, an environmental set point (for Neurotcism, Extraversion, and Agreeableness), and environmental change ($G_\xi + E_\xi + E_\zeta + \lambda_E$) were most suitable to describe the data of the middle-adulthood sample.

Across all personality variables, the exclusion of the inventory method factor did not lead to a decline in fit, $-2LL$ differences ranged between 0.00 and 0.78 ($\Delta -2LL_{\text{crit},p} < .10(1) = 2.71$). Thus, the choice of inventories (NEO-FFI vs. NEO-PI-R) did not affect our results. Testing the equivalence of the degree of genetic set-point variance between age groups did not lead to a decline in fit, too ($\Delta -2LL$ ranged between 0.00 and 1.80; $\Delta -2LL_{\text{crit},p} < .10(1) = 2.71$).

The view to the phenotypic and biometric parameters derived from the best fitting model offers more specific information. Table 5 represents all latent variable variance components derived from the best fitting models (last rows of Table 4) aside from random error (ε). Set-point (ξ) variance in the young subsample was exclusively influenced genetically. In Neuroticism, Extraversion, and Agreeableness, there was an increase in set-point (ξ) variance across age samples as a result of significant environmental set-point variance in the older subsample. Variance due to short- and long-term change ($\tau - \xi$) was smaller in middle adulthood (except Time 3 for Conscientiousness) and solely affected by the environment. The younger subsample, however, showed a larger degree of rank-order change and occasion-specific effects. The decrease of occasion-specific effects and the exclusive genetic set-point variance in young adulthood beyond environmental influences was in line with the genetic set point as well as the strong genetic maturation hypotheses. The increase, however, of environmental set-point variance across age supported the gene–environment transaction hypothesis.

Somewhat surprisingly, a very large proportion of self-report specific (ν_S) variance was found to be stable across time and age subsamples. This stable component showed substantial genetic influences consistently across personality variables and age subsamples, indicating that self-reports include large portions of self-report-specific components (not shared with peer reports) that were long-term stable and basically attributable to genetic factors.

Adding up all genetic and environmental variance components for each method at each measurement occasion corrected for instable method effects and random error, we calculated heritability estimates of 59%–79% (*Mdn* = 68%) for self-reports and 51%–81% (*Mdn* = 67%) for peer reports in the young sample as well as 57%–71% (*Mdn* = 64%) for self-reports and 49%–78% (*Mdn* = 60%) for peer reports in the middle-aged sample. We found the same pattern for true score variance corrected for stable method effects (ν_S and ν_P): Heritability decreased from 68% (51%–85%) in the young sample to 65% (55%–80%) in the middle-aged sample. This indicates, first, a larger heritability for personality reports when corrected for both instable and stable method effects as well as random error and, second, slight decreases of heritability for both self- and peer reports across age, a finding that has already been reported for (uncorrected) self-reports by Viken et al. (1994). Larger heritability coefficients for true score variance components provided support for the genetic maturation hypothesis that postulated strong rater-specific components due to nongenetic method effects. The decrease of heritability, however, was exclusively in line with the gene–environment transaction hypothesis.

Beyond the variance components, we computed latent phenotypic correlations between true scores as well as correlations due to genetic and environmental factors (see Table 6) for each mean age interval.[3] These coefficients reflected continuity corrected for

[3] *Genetic continuity* refers to enduring effects shared by different methods as a function of twins' resemblance. *Environmental continuity* refers to enduring effects that are found in both self-reports and peer ratings but are not shared by twins.

Table 4
Multiple Group Model Fit Statistics

Model	Fit statistic	Variable				
		N	E	O	A	C
Baseline (BL)						
$G_\xi + E_\zeta$	$-2LL(df)$	45692.06 (7025)	42710.40 (7025)	40335.64 (7024)	41697.39 (7024)	42982.37 (7022)
				Complete sample		
(Model 1) BL + $G_\zeta + \lambda_G$	$-2LL(df)$	45687.53 (7015)	42702.35 (7015)	40327.19 (7014)	41685.68 (7014)	42966.42 (7012)
vs. BL	$\Delta -2LL(10)$	4.53	8.05	8.45	11.71	15.95
(Model 2) BL + E_ξ	$-2LL(df)$	45648.96 (7023)	42585.40 (7023)	40299.78 (7022)	41671.25 (7022)	42960.80 (7020)
vs. BL	$\Delta -2LL(2)$	43.10***	125.00***	35.86***	26.14***	21.57***
(Model 3) BL + λ_E	$-2LL(df)$	45623.86 (7021)	42582.76 (7021)	40291.20 (7020)	41667.38 (7020)	42952.79 (7018)
vs. BL	$\Delta -2LL(4)$	68.20***	127.64***	44.44***	30.01***	29.58***
(Model 4) BL + $E_\xi + \lambda_E$	$-2LL(df)$	45620.69 (7019)	42572.29 (7019)	40289.11 (7018)	41660.81 (7018)	42951.39 (7016)
vs. (Model 2) BL + E_ξ	$\Delta -2LL(4)$	28.27***	13.11***	10.67**	10.44**	9.41*
vs. (Model 3) BL + λ_E	$\Delta -2LL(2)$	3.17	10.47***	2.09	6.57**	1.40
				Young adulthood group		
(Model 1) BL + $G_\zeta + \lambda_G$	$-2LL(df)$	45689.98 (7020)	42703.90 (7020)	40333.91 (7019)	41686.32 (7019)	42972.13 (7017)
vs. BL	$\Delta -2LL(5)$	2.08	6.50	1.73	11.07*	10.24*
(Model 2) BL + E_ξ	$-2LL(df)$	45677.97 (7024)	42676.54 (7024)	40328.93 (7023)	41694.95 (7023)	42978.90 (7021)
vs. BL	$\Delta -2LL(1)$	14.09***	33.86***	6.71***	2.44	3.47*
(Model 3) BL + λ_E	$-2LL(df)$	45665.54 (7023)	42666.90 (7023)	40322.91 (7022)	41689.85 (7022)	42972.44 (7020)
vs. BL	$\Delta -2LL(2)$	26.52***	43.50***	12.73***	8.39**	9.93***
(Model 4) BL + $E_\xi + \lambda_E$	$-2LL(df)$	45664.84 (7022)	42665.12 (7022)	40322.91 (7021)	41689.85 (7021)	42972.44 (7019)
vs. (Model 2) BL + E_ξ	$\Delta -2LL(2)$	13.13***	11.42***	6.02**	5.10*	6.46**
vs. (Model 3) BL + λ_E	$\Delta -2LL(1)$	0.70	1.78	0.00	0.00	0.00
				Middle adulthood group		
(Model 1) BL + $G_\zeta + \lambda_G$	$-2LL(df)$	45689.61 (7020)	42708.88 (7020)	40328.95 (7019)	41696.76 (7019)	42976.70 (7017)
vs. BL	$\Delta -2LL(5)$	2.45	1.52	6.69	0.63	5.67
(Model 2) BL + E_ξ	$-2LL(df)$	45664.27 (7024)	42621.44 (7024)	40307.48 (7023)	41674.07 (7023)	42966.76 (7021)
vs. BL	$\Delta -2LL(1)$	27.79***	89.00***	28.16***	23.32***	15.61***
(Model 3) BL + λ_E	$-2LL(df)$	45651.99 (7023)	42627.79 (7023)	40304.30 (7022)	41674.85 (7022)	42963.30 (7020)
vs. BL	$\Delta -2LL(2)$	40.07***	82.61***	31.34***	22.54***	19.07***
(Model 4) BL + $E_\xi + \lambda_E$	$-2LL(df)$	45649.27 (7022)	42619.49 (7022)	40302.67 (7021)	41668.57 (7021)	42962.07 (7019)
vs. (Model 2) BL + E_ξ	$\Delta -2LL(2)$	14.90***	57.05***	4.81*	5.50*	4.69*
vs. (Model 3) BL + λ_E	$\Delta -2LL(1)$	2.72*	8.30***	1.63	6.28**	1.23
Best fitting model	$-2LL(df)$	45621.50 (7020)	42574.31 (7020)	40291.20 (7020)	41648.97 (7014)	42941.82 (7013)
vs. BL	$\Delta -2LL(\Delta df)$	70.56*** (5)	136.09*** (5)	44.44*** (4)	48.42*** (10)	40.55*** (9)

Note. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; G_ξ = genetic set point; E_ζ = occasion-specific environmental influences; $G_\zeta + \lambda_G$ = genetic maturation; E_ξ = environmental set point; λ_E = long-term environmental change; $-2LL$ = -2 log-likelihood; Δ = ratio.
* $p < .10$. ** $p < .05$. *** $p < .01$.

measurement error and method-specific effects. On the basis of the best fitting model, some parameters were zero (e.g., E_ξ). This simplified the formulas whereby, for example, standardized regression coefficients (e.g., λ_{E21} and λ_{E32}) accorded with continuity coefficients (e.g., $r_{E1\text{-}2}$ and $r_{E2\text{-}3}$). Phenotypic continuity increased with time and age and was lower in long-term intervals that reflect rank-order change. Genetic factors influenced long-term stability, whereas genetic effects on rank-order change played only a role in young adulthood (for Agreeableness and Conscientiousness). Environmental factors primarily affected short-term stability and rank-order change in personality. Environmental continuity cumulated with time and age, whereas this increase was consistently larger in young adulthood. Generally, long- and short-term environmental continuity was found to be larger in middle than in young adulthood. The increase of continuity due to environmental factors provided strong support for the gene–environment transaction hypothesis.

Discussion

The primary aim of the present study was to examine the sources of personality rank-order continuity and change. In general, our analyses yielded most support for the gene–environment transaction hypothesis (Caspi et al., 2005), because environmental factors were the primary source of change within and across young and middle adulthood. However, our findings also support aspects of the genetic maturation hypothesis (McCrae et al., 2000) and the genetic set-point hypothesis (Carey, 2002), because set-point variance in young adulthood was exclusively influenced by genetic factors. Occasion-specific effects appear to decrease with age,

Table 5
Best Fitting Models: Phenotypic, Genetic, and Environmental Variance Components

Parameter	Age group	Latent variable	Personality variable N	E	O	A	C
Phenotypic variance components	Young	ξ	17.16	14.38	10.83	7.07	12.28
		$\tau_1 - \xi$	10.58	9.25	4.35	9.45	13.57
		$\tau_2 - \xi$	10.48	7.57	5.10	8.90	7.71
		$\tau_3 - \xi$	16.25	8.93	5.45	5.57	4.30
		ν_S	22.65	13.52	8.78	7.24	13.37
		ν_P	0.22	0.00	0.00	1.98	1.68
	Middle	ξ	21.39	18.59	10.83	9.72	12.28
		$\tau_1 - \xi$	7.22	5.25	3.45	3.03	4.94
		$\tau_2 - \xi$	7.08	3.51	3.83	0.88	5.14
		$\tau_3 - \xi$	8.34	3.42	2.66	2.50	5.93
		ν_S	18.15	12.72	7.71	6.62	11.13
		ν_P	5.47	0.00	0.22	4.17	2.52
Set-point components (%)	Young	G_ξ	17.16 (100)	14.38 (100)	10.83 (100)	7.07 (100)	12.28 (100)
		E_ξ	—	—	—	—	—
	Middle	G_ξ	17.16 (80)	14.38 (77)	10.83 (100)	7.07 (73)	12.28 (100)
		E_ξ	4.23 (20)	4.21 (23)	—	2.65 (27)	—
Trait change and occasional specificity (%)	Young	$G_{\tau 1 - \xi}$	—	—	—	3.99 (42)	5.23 (39)
		$G_{\tau 2 - \xi}$	—	—	—	4.93 (55)	1.46 (19)
		$G_{\tau 3 - \xi}$	—	—	—	2.85 (51)	1.79 (42)
		$E_{\tau 1 - \xi}$	10.58 (100)	9.25 (100)	4.35 (100)	5.46 (58)	8.34 (61)
		$E_{\tau 2 - \xi}$	10.48 (100)	7.57 (100)	5.10 (100)	3.97 (45)	6.25 (81)
		$E_{\tau 3 - \xi}$	16.25 (100)	8.93 (100)	5.45 (100)	2.72 (49)	2.51 (58)
	Middle	$G_{\tau 1 - \xi}$	—	—	—	—	—
		$G_{\tau 2 - \xi}$	—	—	—	—	—
		$G_{\tau 3 - \xi}$	—	—	—	—	—
		$E_{\tau 1 - \xi}$	7.22 (100)	5.25 (100)	3.45 (100)	3.03 (100)	4.94 (100)
		$E_{\tau 2 - \xi}$	7.08 (100)	3.51 (100)	3.83 (100)	0.88 (100)	5.14 (100)
		$E_{\tau 3 - \xi}$	8.34 (100)	3.42 (100)	2.66 (100)	2.50 (100)	5.93 (100)
Self-report specificity (%)	Young	$G_{\nu s}$	15.62 (69)	8.54 (63)	6.59 (75)	3.77 (52)	9.15 (68)
		$E_{\nu s}$	6.03 (31)	4.98 (37)	2.19 (25)	3.47 (48)	4.22 (32)
	Middle	$G_{\nu s}$	12.34 (68)	9.18 (72)	3.17 (41)	3.92 (59)	7.79 (70)
		$E_{\nu s}$	5.81 (32)	3.54 (28)	4.54 (59)	2.70 (41)	3.34 (30)
Peer report specificity (%)	Young	$G_{\nu p}$	0	0	0	0.46 (23)	0.82 (49)
		$E_{\nu p}$	0.22 (100)	0	0	1.52 (77)	0.86 (51)
	Middle	$G_{\nu p}$	0	0	0	1.83 (44)	0
		$E_{\nu p}$	5.47 (100)	0	0.22 (100)	2.34 (56)	2.52 (100)

Note. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; ξ = set point; $\tau - \xi$ = trait change + occasion-specific effects; Indices 1–3 = points of time; ν_S = self-report-specific component; ν_P = peer report-specific component; G = genetic factor; E = environmental factor; dashes represent substitute parameters that were fixed in the best fitted model.

which was a specific deduction from the genetic set-point hypothesis. Moreover, different models fitted the data depending on the age group data and traits that were analyzed. Considering the whole pattern of findings, we thus propose a combined model that integrates relevant assumptions from each of the three theoretical approaches. The resulting conception is in fact very similar to an earlier developmental conception by Scarr and McCartney (1983), proposing that experiences are directed by genotypes.

Cumulative continuity in personality. The increasing phenotypic continuity of personality can be interpreted as a process of developing, committing to, and maintaining an identity (Roberts & Caspi, 2003). Caspi et al. (2005) proposed two mutually supportive life-course dynamics: *social selection* and *social influence*. They assume that individuals select environments that are correlated with their personality traits, which in turn provide experiences that affect personality.

In line with previous behavioral genetic studies on personality development (e.g., McGue et al., 1993; Viken et al., 1994), we found personality stability to be primarily influenced by genetic factors. In particular, we found 13-year continuity of convergent valid true scores in young adulthood to be exclusively influenced by genetic factors. Thereby, the amount of genetic variance did not differ between young and middle-aged adult twins. The exclusive genetic influence on long-term stability in personality traits of young adults supports the assumption of genetic *set-points* (Carey, 2002) or *basic tendencies* (McCrae et al., 2000). Regarding the theory of *social selection* (Caspi et al., 2005), it might be emphasized that young adults select environments that are correlated with their genotypic (not environmental) personality set points. Genotypes could affect emerging personality phenotypes of young individuals possibly both directly and through prompting new experiences (Scarr & McCartney, 1983).

Table 6
Best Fitting Models: Latent Regressions and Convergent Valid Continuity Coefficients

Parameter	Mean age interval	Latent parameter	Variable				
			N	E	O	A	C
Standardized regression parameters	23–29	λ_{G21}	—	—	—	.30	.16
	29–35	λ_{G32}	—	—	—	.29	.07
	23–29	λ_{E21}	.37	.50	.37	.43	.49
	29–35	λ_{E32}	.73	.80	.67	.67	.65
	41–48	λ_{G21}	—	—	—	—	—
	48–55	λ_{G32}	—	—	—	—	—
	41–48	λ_{E21}	.21	.33	.79	.17	.59
	48–55	λ_{E32}	.59	.34	.82	.32	.72
Phenotypic continuity	23–29	$r_{1–2}$	.76	.81	.81	.77	.80
	29–35	$r_{2–3}$	.88	.93	.89	.88	.87
	23–35	$r_{1–3}$	.67	.72	.78	.64	.68
	41–48	$r_{1–2}$	.83	.93	.95	.90	.88
	48–55	$r_{2–3}$	.97	.96	.96	.98	.93
	41–55	$r_{1–3}$	.78	.88	.88	.80	.82
Genetic continuity	23–29	$r_{G1–2}$	1.00	1.00	1.00	.91	.95
	29–35	$r_{G2–3}$	1.00	1.00	1.00	.94	.95
	23–35	$r_{G1–3}$	1.00	1.00	1.00	.79	.82
	41–48	$r_{G1–2}$	1.00	1.00	1.00	1.00	1.00
	48–55	$r_{G2–3}$	1.00	1.00	1.00	1.00	1.00
	41–55	$r_{G1–3}$	1.00	1.00	1.00	1.00	1.00
Environmental continuity	23–29	$r_{E1–2}$	.37	.50	.37	.43	.49
	29–35	$r_{E2–3}$	.73	.80	.67	.67	.65
	23–35	$r_{E1–3}$	.25	.28	.28	.26	.27
	41–48	$r_{E1–2}$	.58	.82	.79	.76	.59
	48–55	$r_{E2–3}$	.94	.89	.82	.95	.72
	41–55	$r_{E1–3}$	.47	.67	.44	.54	.41

Note. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; λ_{21} = regression coefficients of true score 2 on true score 1; λ_{32} = regression coefficients of true score 3 on true score 2; r = latent correlations between different points of Times 1, 2, and 3 corrected for specific effects of self- and peer reports and measurement error; G = genetic component; E = environmental component; dashes represent substitute parameters that were fixed in the best fitted model.

Though long-term stability in young adulthood is not affected by environmental factors in our study, environmental continuity seems to increase (negatively accelerated) with age. In middle adulthood, environmental factors play a significant role in 13-year continuity of Neuroticism, Extraversion, and Agreeableness. The increase of environmental continuity might be the result of a negatively accelerated accumulation of individual-specific *social influences* across the life course, as mentioned by Caspi et al. (2005). However, the process linking environment to personality development has not been studied here. Thus, it remains open whether environmental influences are of a social nature (Caspi et al., 2005) or mediated through biological changes (McCrae & Costa, 2008). The negatively accelerated accumulation might be the result of an increase of personality stabilization (committing to and maintaining an identity) and a decrease of occasion-specific contextual effects (minor effects of or fewer life transitions). Both mechanisms may lead to the well-established cumulative phenotypic continuity of personality (Roberts & DelVecchio, 2000).

The combination of the two mutually supportive life-course dynamics (Caspi et al., 2005) is also known as genotype–environment correlation (Plomin, DeFries, & Loehlin, 1977; Scarr & McCartney, 1983). The personality genotype influences the probability of exposure to certain events (*social selection*) by evoking reactions and seeking out settings as well as modifying and creating situations. In turn, the selected social contexts allocate experiences that affect individuals' development of personality (*social influence*) accumulating across the life span. Thereby, *social selection* directed by personality genotypes should be correlated among relatives because of genetic relatedness (e.g., attending university and majoring in the same field of study), but correlated contexts do not necessarily allocate the same *social influences* on personality (e.g., different study conditions and different fellow students). Maybe, the chance (e.g., allocation to different colleges) plays a minor role in the short but a major role in the long run separating twins' development (Dunn & Plomin, 1990). As a consequence, heritability of personality should rather decrease across adulthood because of the accumulation of specific environmental influences, which is in line with findings on self-reports by Viken et al. (1994). Scarr and McCartney (1983) mentioned that the impetus for certain experiences comes from the genotypes, whereas phenotypes are elaborated and maintained by environments.

Rank-order change. The design of our study allowed us to differentiate between genetic and environmental long-term change and occasion-specific effects. Generally, there was no evidence for genetic change, except for Agreeableness and Conscientiousness in young adulthood. A previous study (Bleidorn et al., 2009), utilizing monomethod data partially from the same twin sample of BiLSAT, found that variance in individual-level change of Agreeableness, Conscientiousness, and Neuroticism was primarily at-

tributable to genetic factors. Our study did not find evidence for genetic change in Neuroticism. One explanation of this divergence might be the sole reliance on self-reports in the previous study. However, it should be noted that the structural equation models in the present study focused exclusively on the relative ordering of individuals and were sensitive to detect the relative change of individuals' ranks. The models were not sensitive to detect genetic and environmental variance in systematic intraindividual-level growth or decline over time, as it was the focus in the previous study. Our analyses were sensitive to detect increase and decrease of relative change over time. Genetic change in Agreeableness and Conscientiousness decreased with age in young adulthood and was not significant in middle adulthood, indicating decreasing effects of genetic maturation on rank-order change across young and middle adulthood.

In contrast, long-term change in personality was consistently found to be predominantly attributable to environmental factors in both young and middle adulthood. Across personality variables, change seems to decrease with age. The higher degree of long-term change in young adulthood may be attributed to more or larger effects of major life transitions in this period of life. In young adulthood, individuals usually have to decide which life goals (e.g., career, family) they primarily want to pursue and how to shape their life course (e.g., vocational training, starting a family). The therewith associated transitions may also contribute to a higher degree of personality trait change. In spite of a higher degree of continuity in later adulthood, long-term stability was still found to be lower than short-term stability, indicating that personality is not fixed in that period of life and change way beyond young adulthood, too.

Occasion-specific effects. According to the definition of true score residuals termed as "state residuals" in the latent state-trait theory (Steyer et al., 1999), occasion-specific effects contain influences of the situation in which the individual's phenotype is measured and (or) effects due to the interaction between person and situation. Unfortunately, in our analyses it was not possible to control for interaction effects. Purcell (2002) showed that interaction between additive genetic factors and specific environments acts like effects of the specific environment in the classical design of twins reared together when interaction is present but not estimated. Thus, environmental occasion-specific effects may be also due to influences of genotype–environment interaction (Plomin et al., 1977): Genetic effects depend on the environments, or, the other way round, environmental effects depend on the genotypes.

Referring to the FFT (McCrae & Costa, 2008), occasion-specific environmental influences may thus be interpreted in terms of *characteristic adaptations* of personality genotypes (*basic tendencies*) that respond to the opportunities and incentives of social contexts. As already mentioned for long-term change, the higher degree of occasion-specific effects in young adulthood may be attributed to larger effects of or easily to more major life transitions in this period of life. Likewise, the revealed occasion-specific genetic effects on Agreeableness and Conscientiousness in young adulthood may be interpreted as *characteristic adaptations* of personality genotypes that are shared between twins reared together, because interaction between additive genetic factors and shared environments acts like additive genetic effects when interaction is present but not estimated (Purcell, 2002).

Self-report-specific stability. A large proportion of method-specific variance in self-reports was stable, whereas the corresponding variance in averaged peer reports was not. In addition, nonshared environmental effects on long-term rank-order stability appear to be first and foremost self-report specific. These results were certainly not the major focus of our study but deserve discussion (Kandler et al., in press). It may indicate that there are stable individual differences in self-report response styles (McCrae & Costa, 2008). In this regard, it should be noted that the self-report method factors across all personality variables and both age subsamples was substantially influenced by genetic factors. Thus, this component may also reflect personality characteristics that are not readily accessible to the peer raters (e.g., motives, self-concept).

This interpretation is in line with Kendler's (2001) two pathways in which genes may be effective: *within-the-skin* and *outside-the-skin*. Genes may affect personality through these two pathways. *Within-the-skin*, genetic expression takes place in internal personality features (e.g., motives and emotions). *Outside-the-skin*, genes can also affect personality-relevant behavior and its social consequences. The *outside-the-skin* pathway is more readily perceivable by peers. Thus, it might be argued that expressions of personality genotypes have to be differentiated into internal (intentions and self-concept) and external (expressed behavior and its social consequences) effects. A recent study found evidence that self- and other raters appear to pay attention to different information cues when judging personality (Hofmann, Gschwender, & Schmitt, 2009). Moreover, observers seem to focus rather on states, whereas self-raters primarily focus on their stable attributes, even when they are instructed to focus on their states. This is in line with our results and the idea that stable method-specific variance in self-reports may reflect valid information on personality not accessible by peers. These results call for future studies addressing this issue. Much remains to be learned about the primary processes of introspective and external personality judgments and perceptions.

Conclusions

In summary, the results of our study led us to formulate main conclusions. First, genetic factors affect rank-order stability in personality directly and possibly through experiences resulting from genotype–environment correlations (Scarr & McCartney, 1983). Second, genetic factors remain stable across adulthood, whereas environmental influences trigger both an increase in phenotypic continuity with age and a decrease in phenotypic continuity with increasing time intervals between assessments. Third, self-report-specific variance is largely stable and genetically influenced, which might reflect internal effects on personality phenotypes that are less perceivable to other persons. Furthermore, we can conclude that much remains to be learned about the primary processes involved in *social selection* and *social influence* as well as the mechanisms underlying gene–environment transaction (Caspi et al., 2005) and genotype–environment correlation affecting personality development. The present work may just serve as a basis in order to continue with research into the processes and mechanisms of personality development over the life span.

References

Bleidorn, W., Kandler, C., Riemann, R., Angleitner, A., & Spinath, F. (2009). Patterns and sources of adult personality development: Growth curve analyses of the NEO-PI-R scales in a longitudinal twin study. *Journal of Personality and Social Psychology, 97,* 142–155.

Blonigen, D. M., Carlson, M. D., Hicks, B. H., Krueger, R. F., & Iacono, W. G. (2008). Stability and change in personality traits from late adolescence to early adulthood: A longitudinal twin study. *Journal of Personality, 76,* 229–266.

Boomsma, D. I., & Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I. simplex models. *Behavior Genetics, 17,* 111–123.

Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI*) [NEO Five-Factor Inventory]. Göttingen, Germany: Hogrefe.

Bouchard, T. J., & Loehlin, J. C. (2001). Genes, evolution, and personality. *Behavior Genetics, 31,* 243–273.

Bratko, D., & Butkovic, A. (2006). Stability of genetic and environmental effects from adolescence to young adulthood: Results of Croatian longitudinal twin study of personality. *Twin Research and Human Genetics, 10,* 151–157.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Carey, G. (2002). *Human genetics for the social sciences.* Thousand Oaks, CA: Sage.

Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry, 12,* 49–66.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56,* 453–484.

Costa, P. T., & McCrae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings of the NEO personality inventory. *Journal of Personality and Social Psychology, 54,* 853–863.

Costa, P. T., & McCrae, R. R. (1989). *NEO Personality Inventory: Manual Form S and Form R.* Odessa, FL: Psychological Assessment Resources.

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

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2006). Statistical power to detect genetic and environmental influences in the presence of data missing at random. *Twin Research and Human Genetics, 10,* 159–167.

Dunn, J., & Plomin, R. (1990). *Separate lives: Why siblings are so different?* New York, NY: Basic Books.

Fraley, C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review, 112,* 60–74.

Hewitt, J. K., Eaves, L. J., Neale, M. C., & Meyer, J. M. (1988). Resolving causes of developmental continuity or "tracking". I. Longitudinal twin studies during growth. *Behavior Genetics, 18,* 133–151.

Hofmann, W., Gschwender, T., & Schmitt, M. (2009). The road to the unconscious self not taken: Discrepancies between self- and observer-inferences about implicit dispositions from nonverbal behavioural cues. *European Journal of Personality, 23,* 343–366.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality, 8,* 149–162.

Johnson, W., McGue, M., & Krueger, R. F. (2005). Personality stability in late adulthood: A behavioral genetic analysis. *Journal of Personality, 73,* 523–551.

Kandler, C., Riemann, R., & Kämpfe, N. (2009). Genetic and environmental mediation between measures of personality and family environment in twins reared together. *Behavior Genetics, 39,* 24–35.

Kandler, C., Riemann, R., Spinath, F. M., & Angleitner, A. (in press). Sources of variance in personality facets: A multiple-rater twin study of self-peer, peer-peer, and self-self (dis-) agreement. *Journal of Personality.*

Keller, M. C., Coventry, W. L., Heath, A. C., & Martin, N. G. (2005). Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. *Behavior Genetics, 35,* 707–721.

Kendler, K. S. (2001). Twin studies of psychiatric illness: An update. *Archives of General Psychiatry, 58,* 1005–1014.

Krueger, R. F., Markon, K. E., & Bouchard, T. J., Jr. (2003). The extended genotype: The heritability of personality accounts for the heritability of recalled family environments in twins reared apart. *Journal of Personality, 71,* 809–834.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of American Statistical Association, 83,* 1198–1202.

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In O. P. John, R. W., Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 159–181). New York, NY: Guilford Press.

McCrae, R. R., Costa, P. T., Ostendorf, F., Angleitner, A., Høebíčková, M., Avia, M. D., . . . Smith, P. B. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology, 78,* 173–186.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60,* 175–215.

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

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. M. (2003). *Mx: Statistical modeling.* VCU Box 900126, Richmond, VA 23298, 6th ed. Department of Psychiatry.

Neale, M. C., & Maes, H. H. M. (2004). *Methodology for genetic studies of twins and families.* Dordrecht, the Netherlands: Kluwer Academic Publishers B. V.

Ostendorf, F., & Angleitner, A. (2004). *NEO-PI-R: NEO-Persönlichkeitsinventar nach Costa und McCrae, revidierte Fassung* [NEO-Personality Inventory revised by Costa and McCrae]. Göttingen, Germany: Hogrefe.

Pederson, N. L., & Reynolds, C. A. (1998). Stability and change in adult personality: Genetic and environmental components. *European Journal of Personality, 12,* 365–386.

Pederson, N. L., & Reynolds, C. A. (2002). Stability and change in adult personality: Genetic and environmental components [Erratum]. *European Journal of Personality, 16,* 77–78.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype–environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84,* 309–322.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5,* 554–571.

Riemann, R., Angleitner, A., & Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins reared together using self- and peer report NEO-FFI scales. *Journal of Personality, 65,* 449–475.

Riemann, R., & Kandler, C. (in press). Construct validation using multitrait-multimethod twin data: The case of a general factor of personality. *European Journal of Personality.*

Roberts, B. W., & Caspi, A. (2003). The cumulative continuity model of personality development: Striking a balance between continuity and change in personality traits across the life course. In U. M. Staudinger & U. Lindenberger (Eds.) *Understanding human development: Dialoges with lifespan psychology* (pp. 183–214). Dorrecht, the Netherlands: Kluwer Academic Publishers.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency

of personality from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126,* 3–25.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype → environment effects. *Child Development, 54,* 424–435.

Spinath, F. M., Wolf, H., Angleitner, A., Borkenau, P., & Riemann, R. (2005). Multimodale Untersuchung von Persönlichkeit und kognitiven Fähigkeiten. Ergebnisse der deutschen Zwillingsstudie BiLSAT und GOSAT [Multimodel investigation of personality and cognitive abilities. Results of the German twin studies BiLAST and GOSAT]. *Zeitschrift für Soziologie in der Erziehung, 25,* 146–161.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality, 13,* 389–408.

Terracciano, A., Costa, P. T., Jr., & McCrae, R. R. (2006). Personality plasticity after age 30. *Personality and Social Psychology Bulletin, 32,* 999–1009.

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.

Yamagata, S., Suzuki, A., Ando, J., Ono, Y., Kijima, N., Yoshimura, K., . . . Jang, K. L. (2006). Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. *Journal of Personality and Social Psychology, 90,* 987–998.

Received May 7, 2009
Revision received February 28, 2010
Accepted March 19, 2010 ■