

Higher-Order Factors of the Big Five in a Multi-Informant Sample

Colin G. DeYoung
Yale University

In a large community sample ($N = 490$), the Big Five were not orthogonal when modeled as latent variables representing the shared variance of reports from 4 different informants. Additionally, the standard higher-order factor structure was present in latent space: Neuroticism (reversed), Agreeableness, and Conscientiousness formed one factor, labeled *Stability*, and Extraversion and Openness/Intellect formed a second factor, labeled *Plasticity*. Comparison of two instruments, the Big Five Inventory and the Mini-Markers, supported the hypotheses that single-adjective rating instruments are likely to yield lower interrater agreement than phrase rating instruments and that lower interrater agreement is associated with weaker correlations among the Big Five and a less coherent higher-order factor structure. In conclusion, an interpretation of the higher-order factors is discussed, including possible neurobiological substrates.

Keywords: Big Five, metatraits, stability, plasticity, higher-order factors

One of the major concerns in personality psychology is the development of a comprehensive model of personality traits, typically conceived as a hierarchy in which correlated lower level traits are grouped together within broader higher level traits. The five-factor model, or Big Five, is a promising candidate (though there is some debate as to whether six- or seven-factor models would be more appropriate; Ashton et al., 2004; Saucier & Goldberg, 2001). The Big Five trait domains—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness/Intellect—have often been conceived as orthogonal factors and the highest, most general level of the hierarchy of personality traits (Costa & McCrae, 1992a, 1992b; Goldberg, 1993). Investigation of correlations among the Big Five, however, has demonstrated that they are not orthogonal (at least as currently measured; Saucier, 2002) and that they possess a stable higher-order factor solution (DeYoung, Peterson, & Higgins, 2002; Digman, 1997; cf. Markon, Krueger, & Watson, 2005). Emotional Stability (Neuroticism reversed), Agreeableness, and Conscientiousness mark a first factor, whereas Extraversion and Openness/Intellect mark a second. Although Digman (1997, p. 1248) gave the higher-order factors, or metatraits, the “provisional” labels α and β , we have suggested that they be labeled *Stability* and *Plasticity* (DeYoung et al., 2002). The shared variance of Neuroticism, Agreeableness, and Conscientiousness appears to reflect the individual’s ability and tendency to maintain stability and avoid disruption in emotional, social, and motivational domains, whereas the shared

variance of Extraversion and Openness/Intellect appears to reflect the ability and tendency to explore and engage flexibly with novelty, in both behavior and cognition (DeYoung et al., 2002; DeYoung, Peterson, & Higgins, 2005).

Many questions remain regarding interpretation and explanation of the metatraits. The value of discussing these issues, however, is contingent on the answer to a more basic question: Are the correlations among the Big Five real? Although Big Five scores routinely show intercorrelations, and the higher-order factors have been demonstrated with a variety of instruments and in both self- and observer ratings, several arguments have been made against the substantive reality of these correlations.

Costa and McCrae (1992b) have argued that correlations among the Big Five are method artifacts, stemming from the idiosyncrasies of individual instruments. This argument is weakened by demonstrations that the Big Five are correlated even when latent variables are derived from single-informant ratings on multiple instruments (e.g., John & Srivastava, 1999; Yik & Russell, 2001). McCrae and Costa (1999) have also argued that the higher-order factors merely reflect biases in personality assessment, along two evaluative dimensions: Positive Valence (PV) and Negative Valence (NV). Their own prior work counters this assertion, however, in that they found that PV and NV were not associated with biased self-reports of the Big Five (McCrae & Costa, 1995). That the two evaluative dimensions are similar to the metatraits in their associations with the Big Five (McCrae & Costa, 1999), but do not seem to be associated with biased personality ratings, suggests instead that very general evaluations may be based on the metatraits.

More recently, Biesanz and West (2004) have argued that correlations among the Big Five are indeed method artifacts, resulting not from the characteristics of individual instruments but from the biases of individual raters. Using confirmatory factor analysis (CFA), these authors found that latent Big Five variables representing the shared variance of self-, peer, and parent reports were uncorrelated, despite the fact that all three

This study was supported by a grant from the Social Sciences and Humanities Research Council of Canada to Jordan B. Peterson and by an Ontario Graduate Scholarship. Thanks go to Lewis R. Goldberg for his generosity in making these data available and to Lena C. Quilty and William A. Cunningham for statistical advice.

Correspondence concerning this article should be addressed to Colin G. DeYoung, Department of Psychology, Yale University, Box, 208205, New Haven, CT, 06520. E-mail: cdeyoung@post.harvard.edu

types of rating showed significant intercorrelations among the Big Five when examined separately. Had these findings been reliable, all discussion of higher-order personality factors above the Big Five would be pointless, except inasmuch as one is interested in systematic biases in personality perception. An obvious flaw in Biesanz and West's argument, however, is that interrater agreement in their sample was quite low. The mean correlation of ratings of the same trait by different informants was only .30 (range = .18–.43). By contrast, Costa and McCrae (1992a) reported cross-informant correlations for the Revised NEO Personality Inventory (NEO PI-R) with a mean of .47 (range = .30–.67). Given low interrater agreement (i.e., low correlations between different informants' ratings of the same trait), the correlations between different informants' ratings of different traits are likely to be affected as well. Poor agreement may reduce not only the magnitude of different-trait, different-informant correlations but also their systematicity. Reductions either in the consistency of the pattern of correlations or in their magnitude will decrease the likelihood that significant correlations will be evident among the Big Five in latent space.¹

The present study involved multi-trait multi-method (MTMM) analyses conducted with a large data set with four informants' ratings of the Big Five, where each informant was treated as a different method. Results for two instruments, the Big Five Inventory (BFI; John & Srivastava, 1999) and the Mini-Markers (Saucier, 1994), were compared. A first hypothesis was that, in a sample with greater interrater agreement, significant correlations would be evident among latent Big Five traits. A second hypothesis was that, given adequate interrater agreement, the metatraits of Stability and Plasticity would be present as higher-order factors at the latent level.

An additional question regarding the metatraits was whether they would be correlated. In his CFAs, Digman (1997) tested only models with the correlation between metatraits fixed at zero. We have found, however, that latent metatraits are fairly strongly correlated; in two samples, the correlations between Stability and Plasticity were .45 and .53 (DeYoung et al., 2002). In reanalysis of Digman's (1997) data, Mutch (2005) found that when Digman's CFA procedure was corrected to account for the fact that his data consisted of correlation rather than covariance matrices, a two-factor solution did not fit the data well. This lack of fit might be attributable to the fact that Digman did not allow the metatraits to correlate. The present study therefore examined whether the metatraits were correlated, both in latent traits representing the shared variance across informants and within ratings by single informants. Correlations between the metatraits in data from single informants may reflect a tendency for informants to describe people positively or negatively—as having socially desirable or undesirable qualities—across all trait dimensions, in which case latent metatraits derived from ratings of multiple informants might be less strongly correlated. These latent metatraits would contain only variance agreed upon by all four informants, which should be more reliably linked to the observable patterns of behavior that constitute the substance of the Big Five, rather than to individual informants' judgments about desirability.

Another hypothesis was formulated in response to the question of why interrater agreement was so low in Biesanz and West's (2004) sample. Whereas many factors could have contributed to poor agreement among informants, one likely possibility presents

itself: The instrument used by Biesanz and West to assess the Big Five was Goldberg's (1992) measure of 100 trait-descriptive adjectives (TDA). Single adjectives are known to produce less consistent personality ratings than longer items that provide more context, because single words are more subject to idiosyncrasies in the interpretation of their meaning (Goldberg & Kilkowski, 1985; John & Srivastava, 1999). One reason to expect greater interrater agreement in the current study is that BFI items incorporate prototypical adjective markers of the Big Five into short phrases (e.g. "is emotionally stable, not easily upset"), with the explicit intent of increasing the consistency of ratings (John & Srivastava, 1999). Because the current study incorporates a single-adjective rating instrument (the Mini-Markers) as well as the BFI, the hypothesis that use of a single-adjective rating instrument will be associated with lower interrater agreement can be tested directly. Assuming that this hypothesis holds true, a related hypothesis is that lower interrater agreement will be associated with decreased correlations among latent Big Five traits.

Although the primary purpose of analyzing ratings from both the BFI and Mini-Markers was to examine the relation of interrater agreement to latent trait correlations across different instruments, the inclusion of multiple instruments also served a second purpose. A model could be fit in which each informant's Big Five ratings were modeled as latent variables with two indicators, with scores from the BFI and Mini-Markers serving as separate indicators. Method effects could then be assessed for the two instruments and for each informant, simultaneously. This allowed an assessment of what is common across the correlational structures of both instruments in addition to ways in which they differ.

Method

Participants

Participants were 490 members of the Eugene–Springfield Community Sample, ranging in age from 18 to 80 years ($M = 51.23$, $SD = 12.62$). Participants were recruited by mail from lists of homeowners and agreed to

¹ Biesanz and West (2004) argued against the possibility that low correlations across informants could be responsible for the lack of correlation among the latent Big Five in their multiple-informant analysis, but their reasoning is questionable: They compare their results for multi-informant ratings to their results (in the same sample) for self-ratings at multiple time points. Latent Big Five variables derived from self-ratings at three different times showed significant correlations and continued to do so even when correlations among the self-ratings were reduced artificially by increasing the variance of the scores while maintaining the same covariances. In their general discussion, Biesanz and West (2004) stated that this procedure rendered the magnitude of the correlations in the self-report analysis "comparable" (Biesanz and West, 2004, p. 869) to that in the multi-informant analysis. Earlier in their article, however, they noted that their procedure "had the effect of reducing correlations among measures by approximately 50%" (p. 860). The mean different-trait, different-time absolute correlation in their self-report analysis was .25, and half of this (.12) is still more than twice as large as the mean different-trait, different-informant absolute correlation in their multi-informant analysis (.05). I would suggest that .05 is not sufficiently comparable in magnitude to .12 to settle the question of whether interrater agreement is responsible for Biesanz and West's failure to find significant correlations among latent Big Five traits.

complete questionnaires, delivered by mail, for pay. Self-reports and ratings of three additional informants on the BFI were available for 483 participants (283 female, 200 male). Self-reports and three additional informant ratings on the Mini-Markers were available for 487 participants (283 female, 204 male). The sample spanned all levels of educational attainment, with an average of 2 years of post-secondary schooling. Most participants were Caucasian American (97%), with 1% or less (for each category) identifying as Hispanic, Asian American, or Native American or not reporting their ethnicity.

Participants were instructed that, in addition to filling out self-report questionnaires, they should distribute additional copies designed for peer ratings to any three people who knew them "very well." These additional 1,470 informants (550 female, 914 male, 6 with no gender reported) ranged in age from 6 to 94 years ($M = 48.17$, $SD = 17.99$). (Because of the possibility that children may provide less reliable personality ratings, analyses were repeated with various cutoffs for age, excluding participants with ratings from informants younger than 10, 13, or 17 years. For all three of these cutoffs, results were extremely similar to those obtained in the full sample; hence, only the latter are reported.) Participants described 2.3% of the additional informants as "significant other," 21.7% as "spouse," 28.0% as "friend," 11.4% as "co-worker," 27.9% as "relative," 1.2% as "acquaintance," and 6.3% as "other." No relationship status was reported for 1.3% of informants. On the whole, these raters felt favorably toward their targets; each responded to a single Likert-scale item asking how much they liked the participant, with possible responses ranging from 1 (*like very much*) to 6 (*greatly dislike*). The mean response was 1.21 ($SD = 0.51$; range = 1–5).

Measures

Questionnaires were sent to and received from participants by mail. Informants (both self and other) rated participants' personalities using the Big Five Inventory (BFI; John & Srivastava, 1999) and the Mini-Markers (Saucier, 1994). Both scales are well-validated as measures of the Big Five. The BFI consists of 44 descriptive phrases, with each trait indicated by 8 to 10 items. The Mini-Markers consist of 40 adjectives, with each trait indicated by 8 items. The Mini-Markers were created by taking a subset of the adjectives from the TDA, eliminating many difficult or unusual words; therefore, this measure seems likely to produce more reliable ratings than the TDA; indeed, Saucier (1994) found that the Mini-Markers had a higher mean inter-item correlation within each trait scale than did the TDA. All items were rated for accuracy by informants on a 5-point Likert scale. Trait scores were calculated as the mean item score.² Each of the three peer ratings for each participant was assigned randomly to one of three groups.

Analyses

Following examination of the MTMM correlation matrices (Table 1) to assess inter-rater agreement, the seven MTMM models described below were fitted with confirmatory factor analysis (CFA).³ Each informant was treated as a different method. The best fitting model was retained, and statistical comparisons of differences in fit were made to test the performance of models specifying orthogonality over models allowing correlated traits.

1. Correlated traits, no methods (CTNM): Models five latent trait factors and their correlations but does not assume or model any method effects.

2. Correlated traits, correlated uniquenesses (CTCU; see Figure 1A): Models method effects as correlations among the five uniquenesses for each informant. In CFA, a uniqueness represents the variance in an observed variable not explained by latent variables. No assumptions are made regarding the dimensionality of the method effects.

3. Orthogonal traits, correlated uniquenesses (OTCU): Models method effects identically to the CTCU model but assumes that the five latent trait

factors are uncorrelated. The difference in fit between CTCU and OTCU models provides a statistical test of the orthogonality of the Big Five.

4. Correlated traits, correlated methods (CTCM): Unlike the CTCU model, this model assumes that a single latent factor underlies each method effect. It also allows the latent method factors to be correlated across informants.

5. Correlated traits, orthogonal methods (CTOM; see Figure 1B): Like the CTCM model but assumes that the method factors are uncorrelated. This model is nested under the CTCU model and comparison of these two models permits a test of whether the method effects are unidimensional.

6. Orthogonal traits, correlated methods (OTCM): This model is nested under the CTCM model, being identical to it except for the assumption that the latent traits are uncorrelated.

7. Orthogonal traits, orthogonal methods (OTOM): Assumes uncorrelated latent trait and method factors. The OTOM model is nested under the CTCU, OTCU, CTCM, CTOM, and OTCM models.

All models were analyzed with Amos 5.0 (Arbuckle, 2003) with maximum likelihood estimation based on the full covariance matrices. CFAs of higher-order factor structure follow the investigation of trait correlations.

Results

Interrater Agreement

Correlations between traits as assessed by the BFI and the Mini-Markers within raters (same trait, same informant, different instrument) were quite high (mean $r = .82$; range = .73–.90), indicating that the two scales assess the Big Five very similarly. Nonetheless, as predicted, interrater agreement (correlations for same trait, different informant) was significantly higher for the BFI (mean $r = .41$, $SD = .08$, range = .29–.57) than for the Mini-Markers (mean $r = .36$, $SD = .08$, range = .25–.53), $F(1, 29) = 22.87$, $p < .001$.

Model Fit for Multi-Trait Multi-Method Confirmatory Factor Analyses

Table 2 presents fit indices for the CFAs of the seven models described above, for both the BFI and Mini-Markers. In addition to the chi-square test for significant discrepancies between the predicted and observed covariance matrices, the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) are presented. CFI values over .95 are considered to indicate good fit. RMSEA values less than .08 indicate acceptable fit, whereas

² Factor scores were examined as an alternative to mean item scores. Five factors were extracted from item-level data using principal axis factoring with direct oblimin rotation ($\delta = 0$). Correlations among the Big Five, in both single-informant and multi-informant analyses, remained very similar with this method, and interrater agreement was unchanged. An orthogonal rotation (varimax) produced factor scores with slightly higher mean interrater agreement (.44 instead of .41 for the BFI; .37 instead of .36 for the Mini-Markers), but at the cost of explaining less variance in the items and artificially preventing any test of the hypothesis in question (i.e., that the Big Five are correlated). An oblique rotation is the appropriate test for whether the underlying factors are correlated in single-informant ratings.

³ Models fit with three item-packets as indicators for each of the Big Five, in order to create latent trait variables for each informant (thereby adding a lower level of latent variables to the model) produced nearly identical results and are not reported because of their additional complexity.

Table 1
Multi-Trait Multi-Method Correlation Matrices for the Big Five Inventory and the Mini-Markers

Informant/factor	Self-report					Peer 1 report					Peer 2 report					Peer 3 report				
	E	A	C	N	O	E	A	C	N	O	E	A	C	N	O	E	A	C	N	O
Self																				
Extraversion	—	.19	.12	-.04	.19	.53	.00	-.01	.07	.07	.53	.00	-.02	.05	.05	.51	-.04	-.05	.11	-.01
Agreeableness	.15	—	.15	-.35	.09	.12	.32	-.01	-.15	.01	.05	.30	-.04	-.06	.08	.05	.30	.02	-.08	.10
Conscientiousness	.21	.24	—	-.17	.08	.08	-.05	.39	-.10	-.05	.01	-.07	.34	.03	-.03	-.03	-.06	.33	.00	-.03
Neuroticism	-.16	-.36	-.31	—	-.05	-.03	-.15	-.08	.33	-.07	.01	-.18	-.01	.27	-.05	-.02	-.20	-.12	.28	-.06
Openness/Intellect	.25	.06	.09	-.08	—	.01	-.10	-.10	.10	.35	.09	-.02	-.06	.05	.39	.02	-.08	-.06	.08	.38
Peer 1																				
Extraversion	.57	.10	.11	-.07	.12	—	.08	.13	-.08	.15	.47	-.01	.03	.05	.03	.40	.00	-.04	.06	-.03
Agreeableness	-.01	.38	-.01	-.16	-.08	.14	—	.26	-.51	.23	.01	.28	.07	-.11	.06	.04	.30	.11	-.17	.06
Conscientiousness	.00	.03	.35	-.12	-.15	.16	.30	—	-.32	.27	.01	.02	.43	-.07	.02	.02	.04	.38	-.05	-.02
Neuroticism	.01	-.15	-.07	.40	.08	-.15	-.54	-.38	—	-.11	.03	-.18	-.16	.31	-.02	.02	-.21	-.16	.31	-.08
Openness/Intellect	.12	.03	-.04	-.07	.44	.28	.25	.18	-.16	—	.09	.04	.06	-.04	.36	.07	-.01	.10	.05	.29
Peer 2																				
Extraversion	.56	.01	.02	-.07	.15	.49	.06	-.01	-.05	.14	—	.07	.14	-.12	.19	.47	-.03	.01	.05	.06
Agreeableness	-.03	.29	-.08	-.15	-.06	.01	.35	.06	-.22	.00	.12	—	.23	-.55	.24	-.04	.25	.08	-.19	.05
Conscientiousness	.00	.00	.33	-.08	-.09	.03	.11	.39	-.16	.01	.15	.35	—	-.28	.25	-.06	.09	.41	-.13	.03
Neuroticism	-.03	-.09	-.07	.37	.01	-.01	-.15	-.12	.38	.04	-.24	-.53	-.41	—	-.11	.08	-.16	-.11	.29	-.02
Openness/Intellect	.15	.01	-.09	.02	.48	.10	.02	-.06	.01	.41	.35	.22	.19	-.18	—	.05	.03	.08	-.02	.41
Peer 3																				
Extraversion	.54	.06	.04	-.08	.08	.45	.04	.03	-.02	.10	.52	.02	.04	-.02	.17	—	.06	.06	.02	.08
Agreeableness	.01	.32	.00	-.18	-.10	.02	.36	.07	-.24	-.03	.01	.34	.13	-.22	.02	.16	—	.32	-.59	.26
Conscientiousness	.00	.06	.31	-.14	-.09	.05	.17	.35	-.23	.01	.10	.37	-.14	-.02	.21	.43	—	-.30	.26	
Neuroticism	.05	-.10	-.04	.42	.08	-.00	-.24	-.12	.42	.01	-.04	-.22	-.13	.40	-.04	-.14	-.59	-.44	—	-.07
Openness/Intellect	.10	.03	-.03	-.05	.52	.07	.03	-.09	-.05	.41	.13	.05	-.02	-.04	.44	.28	.23	.24	-.19	—
Big Five Inventory																				
<i>M</i>	3.34	4.07	4.08	2.56	3.69	3.63	4.11	4.19	2.59	3.72	3.68	4.07	4.19	2.64	3.71	3.69	4.11	4.25	2.57	3.72
<i>SD</i>	0.80	0.57	0.60	0.78	0.70	0.78	0.76	0.67	0.91	0.72	0.82	0.74	0.68	0.88	0.71	0.78	0.80	0.69	0.90	0.71
Alpha	.87	.80	.84	.85	.85	.86	.87	.86	.87	.86	.86	.87	.85	.87	.85	.83	.90	.86	.88	.84
Mini-Markers																				
<i>M</i>	3.46	4.27	4.06	2.37	3.81	3.66	4.26	4.05	2.49	3.90	3.71	4.21	4.08	2.49	3.91	3.70	4.23	4.13	2.44	3.93
<i>SD</i>	0.78	0.54	0.66	0.72	0.65	0.74	0.71	0.77	0.82	0.65	0.76	0.66	0.74	0.81	0.66	0.76	0.72	0.73	0.82	0.64
Alpha	.85	.81	.86	.81	.83	.81	.83	.85	.82	.82	.81	.84	.85	.82	.82	.81	.87	.85	.82	.78

Note. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Neuroticism; O = Openness/Intellect. $N = 483$ for the Big Five Inventory, and $N = 487$ for the Mini-Markers. Correlations greater than $|.08|$ are significant at $p < .05$. Correlations within informants and correlations of the same trait between different informants are in boldface. Correlations of different traits between different informants are in plain text. Big Five Inventory correlations are below the diagonal, and Mini-Markers correlations are above the diagonal.

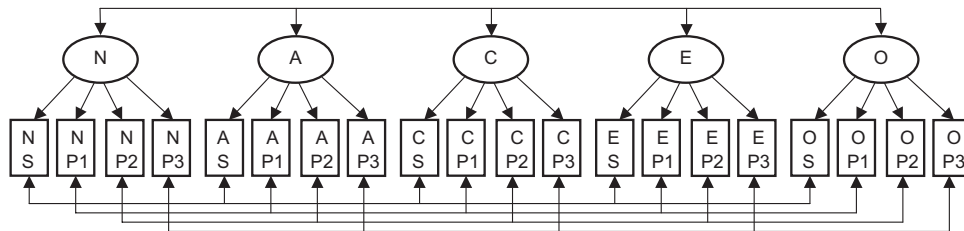
values less than .05 indicate close fit (Kline, 2005). For both the BFI and the Mini-Markers, the CTCU model (Figure 1A) was clearly the best, being the only model with a nonsignificant or nearly nonsignificant chi-square value, which indicates that the covariance matrix predicted by the model does not differ substantially from the observed matrix. (Because the chi-square value is sensitive to sample size, use of a large sample will often cause even good models to differ significantly from the observed data at the traditional significance level of $p < .05$; Kline, 2005.) In addition to having the lowest chi-square, the CTCU model also had the highest CFI values and the lowest RMSEAs.

Correlations Among Latent Traits

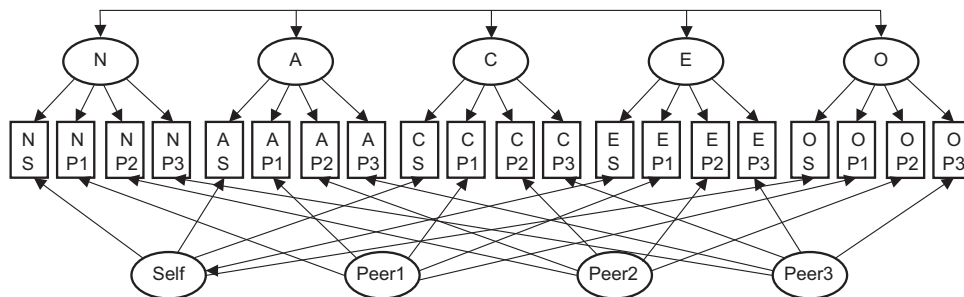
The overall orthogonality of the Big Five was tested by chi-square difference tests comparing the CTCU and OTCU models. If the Big Five were orthogonal, the fit of these two models should not differ significantly. Orthogonality was rejected for both instruments: for the BFI, $\chi^2_{\text{difference}}(10, N = 483) = 90.37, p < .00001$; for the Mini-Markers, $\chi^2_{\text{difference}}(10, N = 487) = 65.78, p < .00001$. Thus, even when the effects of specific informants on

ratings were removed, by creating latent variables representing variance shared across all informants, the Big Five remain significantly intercorrelated.

Table 3 presents the parameter estimates for the CTCU model for both the BFI and the Mini-Markers. As predicted, some of the correlations among the Big Five were significant. Also as predicted, the correlations were generally weaker for ratings obtained with the Mini-Markers, a single-adjective-rating instrument, than for ratings obtained with the BFI, a phrase-rating instrument. The average absolute correlation among the Big Five for the BFI was .15, whereas for the Mini-Markers a correlation of .11 was obtained. (Both of these were higher than the average absolute correlation of .09 reported by Biesanz & West, 2004). The pattern of correlations, particularly for the BFI, appears consistent with the standard higher-order factor model (DeYoung et al., 2002; Digman, 1997), as Neuroticism, Agreeableness, and Conscientiousness are correlated, and Extraversion and Openness/Intellect are correlated. A formal test of this model is presented following analysis of the correlations among uniquenesses in the CTCU model.



A. Correlated Traits, Correlated Uniquenesses (CTCU)



B. Correlated Traits, Orthogonal Methods (CTOM)

Figure 1. Two multi-trait, multi-method confirmatory factor-analytic models of Big Five traits (N = Neuroticism; A = Agreeableness; C = Conscientiousness; E = Extraversion; O = Openness/Intellect) based on ratings from four informants: S = self-ratings; P = peer ratings.

Table 2
Model Fit Indices for Multi-Trait Multi-Method Confirmatory Factor Analyses of the Big Five Inventory and the Mini-Markers

Model	χ^2	df	CFI	RMSEA ^a
Big Five Inventory (N = 483)				
CTNM	1577.04**	160	.61	.136 (.130-.142)
CTCU	136.01	120	.99	.017 (.000-.029)
OTCU	226.38**	130	.97	.039 (.031-.048)
CTCM	244.17**	134	.97	.041 (.033-.049)
CTOM	262.35**	140	.97	.043 (.035-.050)
OTCM	312.53**	144	.95	.049 (.042-.057)
OTOM	363.78**	150	.94	.054 (.047-.062)
Mini-Markers (N = 487)				
CTNM	1194.59**	160	.63	.115 (.109-.122)
CTCU	149.93*	120	.99	.023 (.007-.033)
OTCU	215.71**	130	.97	.037 (.028-.045)
CTCM	294.41**	134	.94	.050 (.042-.057)
CTOM	318.93**	140	.94	.051 (.044-.059)
OTCM	367.19**	144	.92	.056 (.049-.064)
OTOM	399.48**	150	.91	.059 (.052-.065)

Note. CTNM = correlated traits, no methods; CTCU = correlated traits, correlated uniquenesses; OTCU = orthogonal traits, correlated uniquenesses; CTCM = correlated traits, correlated methods; CTOM = correlated traits, orthogonal methods; OTCM = orthogonal traits, correlated methods; OTOM = orthogonal traits, orthogonal methods; CFI = comparative fit index; RMSEA = root mean square error of approximation. ^a 90% confidence intervals are presented in parentheses. * $p < .05$. ** $p < .001$.

Uniquenesses

As seen in Table 3, correlations among uniquenesses were not only larger in magnitude than correlations among the latent Big Five, they were also larger in magnitude than the same-informant, different-trait correlations shown in boldface in Table 1. This finding indicates that after extracting the variance shared among raters, each individual rater's leftover variance is fairly consistent, which is to say that, relative to the other raters, he or she consistently rated the target as having more desirable or undesirable qualities across all traits. It is hardly surprising that raters' general impressions of the desirability or undesirability of targets' personalities should influence their ratings on all trait dimensions. If the correlations among uniquenesses were due exclusively to such a general bias, however, the method effects associated with individual raters would be unidimensional, and the CTOM model (Figure 1B) would fit as well as the CTCU model (Figure 1A). This was not the case, as chi-square difference tests indicated that the CTCU model was significantly better in fit than the CTOM model: for the BFI, $\chi^2_{\text{difference}}(20, N = 483) = 126.34, p < .00001$; for the Mini-Markers: $\chi^2_{\text{difference}}(20, N = 487) = 169.00, p < .00001$.

Exploratory factor analyses were therefore conducted to examine the dimensional structure of the correlations among uniquenesses. The results of these analyses may be informative regarding the biases present in individual ratings. If the usual two-factor structure were present in the uniquenesses but not in the latent traits, this would suggest that raters' biases are inaccurate and probably entirely responsible for the higher-order factor solution reported in the past (DeYoung et al., 2002; Digman, 1997). If, however, the same two-factor structure were found both in the uniquenesses and at the trait level, this would suggest that people

Table 3
 Parameter Estimates for the Correlated Traits Correlated Uniquenesses Model for the Big Five Inventory and the Mini-Markers
 Shown in Figure 1A

Informant/factor	Big Five Inventory (N = 483)					Mini-Markers (N = 487)				
	N	A	C	E	O	N	A	C	E	O
Factor loadings										
Self-report	.69**	.63**	.61**	.79**	.78**	.58**	.64**	.64**	.78**	.67**
Peer 1	.61**	.60**	.59**	.69**	.61**	.61**	.55**	.62**	.65**	.54**
Peer 2	.57**	.54**	.61**	.71**	.62**	.47**	.49**	.62**	.70**	.61**
Peer 3	.60**	.52**	.53**	.69**	.67**	.46**	.46**	.56**	.67**	.59**
Correlations										
Latent factor										
Neuroticism	—					—				
Agreeableness	-.46**	—				-.49**	—			
Conscientiousness	-.29**	.12†	—			-.22**	.02	—		
Extraversion	-.06	.05	.05	—		.09	.07	.00	—	
Openness/Intellect	.02	-.04	-.16*	.24**	—	-.01	.06	-.04	.10†	—
Uniquenesses										
Self-report										
Neuroticism	—					—				
Agreeableness	-.41**	—				-.31**	—			
Conscientiousness	-.40**	.41**	—			-.22**	.35**	—		
Extraversion	-.25**	.24**	.34**	—		-.11†	.26**	.23**	—	
Openness/Intellect	-.18**	.17**	.34**	.30**	—	-.09	.16*	.26**	.31**	—
Peer 1 report										
Neuroticism	—					—				
Agreeableness	-.56**	—				-.52**	—			
Conscientiousness	-.38**	.35**	—			-.32**	.35**	—		
Extraversion	-.23**	.18**	.21**	—		-.17**	.07	.17**	—	
Openness/Intellect	-.29**	.41**	.37**	.33**	—	-.16**	.33**	.41**	.18**	—
Peer 2 report										
Neuroticism	—					—				
Agreeableness	-.54**	—				-.56**	—			
Conscientiousness	-.48**	.47**	—			-.32**	.32**	—		
Extraversion	-.35**	.21**	.26**	—		-.27**	.13*	.25**	—	
Openness/Intellect	-.29**	.32**	.39**	.35**	—	-.14**	.29**	.40**	.21**	—
Peer 3 report										
Neuroticism	—					—				
Agreeableness	-.62**	—				-.59**	—			
Conscientiousness	-.48**	.48**	—			-.30**	.38**	—		
Extraversion	-.24**	.24**	.30**	—		-.06	.11*	.16**	—	
Openness/Intellect	-.31**	.39**	.45**	.35**	—	-.09†	.34**	.36**	.12*	—

Note. N = Neuroticism; A = Agreeableness; C = Conscientiousness; E = Extraversion; O = Openness/Intellect.

† $p < .10$. * $p < .05$. ** $p < .01$.

have a reasonably accurate (though possibly implicit) expectation regarding how traits covary and also that this expectation leads them to exaggerate trait correlations when they rate their own or others' personalities. Finally, if the factor structure of the uniquenesses and the latent traits were dissimilar, this would suggest that people have inaccurate expectations about which traits vary together.

Maximum likelihood estimation was used for these exploratory factor analyses because it provides a significance test that can be used to evaluate the number of factors necessary to capture the structure of the data. An oblique rotation (direct oblimin, $\delta = 0$) was used to allow for the possibility of correlated factors.

For the BFI, a two-factor solution fit adequately for three of the four sets of uniquenesses, all $\chi^2(1, N = 483) < 2.96, p > .05$. The factor structure for self-rating and peer₁ rating uniquenesses was

the same as in the standard two-factor solution: Neuroticism (reversed), Agreeableness, and Conscientiousness marked the first factor, and Extraversion and Openness/Intellect marked the second. For the peer₃ ratings, Conscientiousness loaded almost equally on both factors. For the peer₂ ratings, the two-factor solution was significantly different from the observed data, $\chi^2(1, N = 483) = 13.46, p < .001$. Three factors were therefore extracted, with principal axis factoring. The first two factors resembled the standard higher-order factors, whereas the third factor was marked primarily by Openness/Intellect. In all four cases, the first two factors were strongly correlated ($r_s > .50$).

For the Mini-Markers, a two-factor solution fit adequately for two of the four sets of uniquenesses, both $\chi^2(1, N = 487) < 1.92, p > .05$. The self-rating uniquenesses showed the standard two-factor solution. In the peer₃ ratings, Neuroticism and Agreeable-

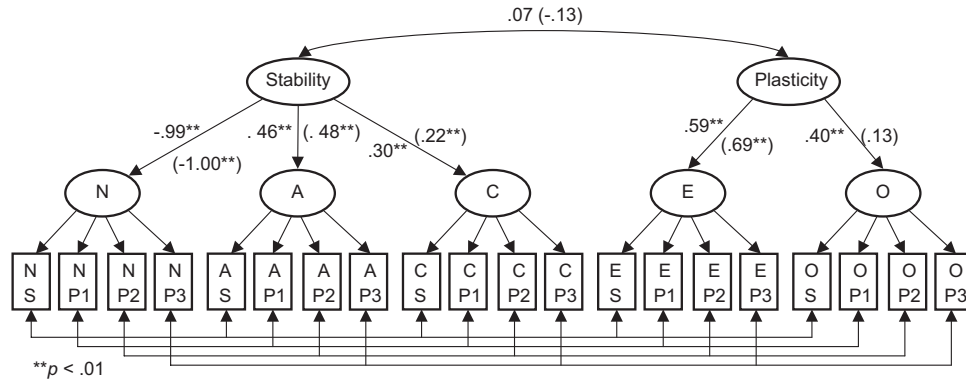


Figure 2. Higher-order factors of the Big Five traits (N = Neuroticism; A = Agreeableness; C = Conscientiousness; E = Extraversion; O = Openness/Intellect), based on ratings from four informants (S = self-ratings; P = peer ratings), with parameter estimates for the Big Five Inventory and the Mini-Markers (estimates for the Mini-Markers are in parentheses). See text for indices of fit and Table 3 for parameter estimates for the measurement model.

ness marked the first factor, Conscientiousness and Openness/Intellect marked the second, and Extraversion loaded weakly on the second factor and not at all on the first. Because the peer₁ and peer₂ ratings were not adequately described by a two-factor solution, both $\chi^2(1, N = 487) > 6.69, p < .05$, three factors were extracted. In both cases, the first factor was marked by Neuroticism (reversed) and Agreeableness, the second by Openness/Intellect and Conscientiousness, and the third by Neuroticism (reversed) and Extraversion. For all four informants, all factors were moderately intercorrelated, with correlations ranging from .26 to .42.

Higher-Order Factors of the Big Five

Confirmatory factor analysis was used to test the hypothesis that the Big Five would show the usual higher-order factor structure, in latent space. Figure 2 depicts a hierarchical model with latent Stability and Plasticity variables above the latent Big Five, with parameter estimates for the higher-order factor solution (see Table 3 for parameter estimates for the measurement model).⁴ For the BFI, this model fit extremely well, $\chi^2(125, N = 483) = 145.11, p = .11$; CFI = .99; RMSEA = .018. Because this model was not nested under the CTCU model, the two could not be compared with the chi-square difference test; however, Akaike's information criterion (AIC) can be used to compare non-nested models, with lower AIC values indicating better fit (Kline, 2005). The higher-order factor model did fit slightly better: for CTCU, AIC = 316.01; for the higher-order factors, AIC = 315.11. The correlation between Stability and Plasticity was nonsignificant, and the fit of the model did not change significantly when the correlation was constrained to zero, $\chi^2(126, N = 483) = 145.59, p = .11$; CFI = .99; RMSEA = .018; $\chi^2_{\text{difference}}(1, N = 483) = 0.48, p = .49$.

The higher-order factor model also fit reasonably well for the Mini-Markers, $\chi^2(126, N = 487) = 157.43, p < .05$; CFI = .99; RMSEA = .023. (The error variance associated with latent Neuroticism was constrained to be non-negative.⁵) Again, AIC values indicated that this model was preferable to the CTCU model: for CTCU, AIC = 329.93; for higher-order factors, AIC = 325.43. Nonetheless, the model was not entirely unproblematic; Openness/

Intellect did not load significantly on Plasticity because of the fact that the correlation between Extraversion and Openness/Intellect was attenuated for the Mini-Markers relative to the BFI (Table 3). As with the BFI, the correlation between Stability and Plasticity was not significant, and the fit of the model did not change significantly when the correlation was constrained to zero, $\chi^2(127, N = 487) = 159.20, p = .03$; CFI = .99; RMSEA = .023; $\chi^2_{\text{difference}}(1, N = 487) = 1.77, p = .18$.

As a test of whether the pattern of correlations among the latent Big Five was significantly multidimensional, the model in Figure 2 was also fitted with the correlation between Stability and Plasticity fixed at unity (1.00). This strategy created a model that is nested under the standard higher-order factor model but is equivalent to a model with only a single higher-order factor marked by all five latent Big Five traits. The model fit well for both instruments: for BFI, $\chi^2(126, N = 483) = 165.77, p = .01$; CFI = .99; RMSEA = .026; for Mini-Markers, $\chi^2(127, N = 487) = 171.02, p < .01$; CFI = .98; RMSEA = .027. However, chi-square difference tests indicated that it did not fit as well as the two-factor model: for BFI, $\chi^2_{\text{difference}}(1, N = 483) = 20.66, p < .001$; for Mini-Markers,

⁴ Because a model containing a latent variable with only two indicators is empirically under-identified if that latent variable is not correlated with another latent variable (Kline, 2005), the unstandardized error variances for the latent Extraversion and Openness/Intellect variables were constrained to be equal, to allow identification of the model.

⁵ This error variance was constrained because without constraint it became negative at some point while the model was fitted. Although negative error variances have sometimes been considered evidence of possible model misspecification, Monte Carlo studies lead to the conclusion that "researchers should not use negative error variance estimates as an indicator of model misspecification" (Chen, Bollen, Paxton, Curran, & Kirby, 2001, p. 501). In the present analyses, the tendency of the error variance for latent Neuroticism to become slightly negative appears to be due to the fact that the value of the loading of Neuroticism on Stability is very near -1.00 (as evidenced by the weight of $-.99$ seen in the model for the BFI, in which the error variance for Neuroticism did not need to be constrained). If the estimate of a loading goes over 1.00 , the associated error variance will become negative.

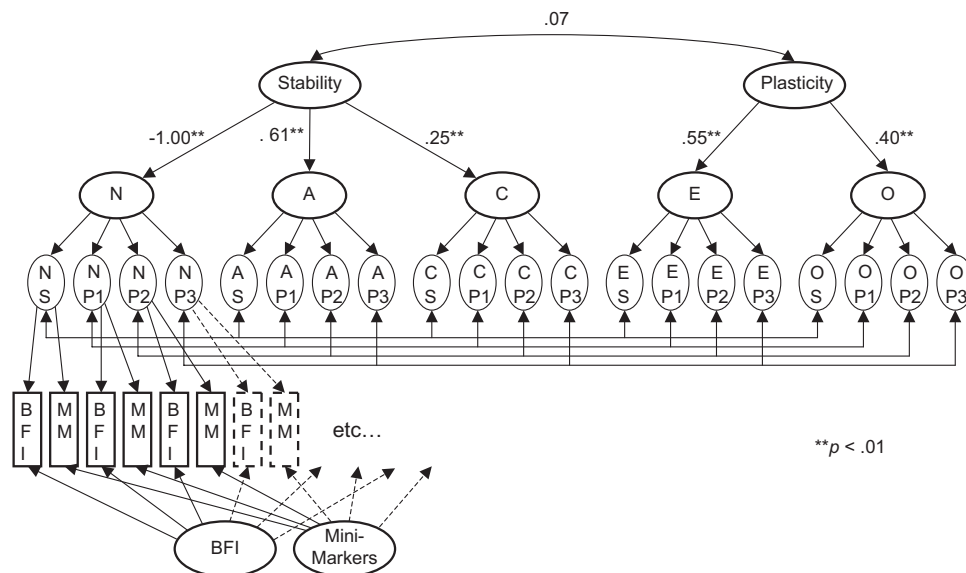


Figure 3. Higher-order factor model for multi-trait, multi-informant, multi-instrument confirmatory factor analysis, based on ratings from four informants (S = self-ratings; P = peer ratings). For clarity of illustration, the full measurement model is not shown. See text for indices of fit and Table 4 for additional parameter estimates. N = Neuroticism; A = Agreeableness; C = Conscientiousness; E = Extraversion; O = Openness/Intellect; BFI = Big Five Inventory; MM = Mini-Markers.

$\chi^2_{\text{difference}}(1, N = 487) = 13.59, p < .001$. The pattern of correlations among the latent Big Five, therefore, is not unidimensional, providing further support for the two-factor model.

Multi-Trait, Multi-Informant, Multi-Instrument Confirmatory Factor Analysis

For the Mini-Markers, correlations among the latent Big Five were reduced relative to the BFI, as predicted, and the standard Plasticity factor marked by both Extraversion and Openness/Intellect was absent. It was therefore of interest to determine whether both metatraits were present if the BFI and Mini-Markers were analyzed simultaneously. Such an analysis would allow some degree of control for method effects specific to each instrument. A model was therefore fitted in which Big Five ratings by each informant were modeled as latent variables with BFI and Mini-Markers scores as separate indicators (Figure 3). Because the CTCU model fit best in the previous analyses, correlated uniquenesses were used at the latent level of individual informant ratings. Two method factors representing variance unique to the BFI and Mini-Markers were included, with each marked by 20 observed variables. (More complex breakdowns of the instrument effects—by trait, for example—were not possible because, in conjunction with the correlated uniquenesses for each informant, they created unidentified models.)

The model fit very well, $\chi^2(636, N = 481) = 1,208.55, p < .01$; CFI = .96, RMSEA = .043.⁶ (Error variance associated with latent Neuroticism was constrained to be non-negative.) Figure 3 displays the factor loadings on the metatraits, and Table 4 displays the loadings of latent traits for each informant on the latent Big Five and loadings of each instrument on each factor for each informant. The loading of Conscientiousness on Stability was low

but significant, and both Extraversion and Openness/Intellect loaded significantly on Plasticity. These results suggest that the absence of the standard Plasticity factor in the model for the Mini-Markers above was due to method variance specific to that instrument. When the Big Five were modeled by the shared variance across both instruments, both metatraits were evident.

Notably, there were no significant loadings of observed variables on the two latent variables representing method effects associated with the different instruments. However, when the model was fitted without the instrument effects factors, the fit of the model was significantly worsened: $\chi^2(676, N = 481) = 1467.32, p < .01$; CFI = .95, RMSEA = .049; $\chi^2_{\text{difference}}(40, N = 481) = 258.77, p < .00001$. The instrument effects factors were therefore retained. Notably, several loadings for the Mini-Markers method factor approached significance ($p < .10$), with loadings ranging between .13 and .29; these included loadings for all four ratings of Openness/Intellect, the three peer ratings of Conscientiousness, and two of the peer ratings of Agreeableness. A pattern suggesting stronger method effects for the Mini-Markers than for the BFI is consistent with the differences in higher-order factor structures for the two instruments when analyzed separately.

Correlations Between Stability and Plasticity

As seen in the models above, once the variance associated with specific informants was removed, Stability and Plasticity were uncorrelated in latent space. Previous work with self-ratings has found substantial correlations between the metatraits (DeYoung et

⁶ The full covariance matrix used to fit this model is available from the author on request.

Table 4
Latent Big Five Factor Loadings for the Multi-Trait,
Multi-Instrument, Multi-Informant Model Shown in Figure 3

Informant/measure	Latent factors				
	N	A	C	E	O
Multi-informant Big Five					
Self-report	.69	.67	.66	.81	.79
Peer 1	.63	.62	.66	.72	.65
Peer 2	.54	.55	.66	.75	.65
Peer 3	.59	.54	.56	.70	.69
Single-informant Big Five					
Self-report					
Big Five Inventory	.83	.96	.96	.97	.98
Mini-Markers	.90	.81	.87	.93	.83
Peer 1 report					
Big Five Inventory	.86	.97	.90	.97	.95
Mini-Markers	.90	.85	.91	.91	.79
Peer 2 report					
Big Five Inventory	.85	.97	.95	.96	.95
Mini-Markers	.89	.86	.88	.93	.83
Peer 3 report					
Big Five Inventory	.86	.97	.98	.99	.99
Mini-Markers	.88	.88	.86	.90	.78

Note. N = 481. N = Neuroticism; A = Agreeableness; C = Conscientiousness; E = Extraversion; O = Openness/Intellect. All loadings were significant at $p < .01$. The loadings in this table correspond to the variables in Figure 3 as follows: multi-informant Big Five loadings are those of the latent (oval) Big Five variables Neuroticism self-report (NS), Neuroticism Peer 1 (NP1), Neuroticism Peer 2 (NP2), and so forth, on the latent Big Five variables Neuroticism (N), Agreeableness (A), Conscientiousness (C), and so forth, whereas single-informant Big Five loadings are those of the observed (rectangular) Big Five variables, labeled BFI (Big Five Inventory) and MM (Mini-Markers), on the latent variables NS, NP1, NP2, and so forth.

al., 2002), but this may have been the result of generally inflated correlations within individual informants' ratings as a result of a bias toward describing oneself as having uniformly desirable or undesirable traits and/or to more specific biases regarding how traits are assumed to covary. The present study allowed further examination of this issue, by fitting CFA models of the higher-order factors with ratings from single informants (Figure 4) and comparing the resulting correlations between Stability and Plasticity with those obtained in the MTMM analysis above. Significant correlations would suggest that the biases of individual raters are responsible for correlations between the metatraits. As shown in Table 5, Stability and Plasticity were significantly correlated in ratings by all four sets of informants, with either instrument.

Table 5 shows the parameter estimates and fit indices obtained for the model depicted in Figure 4 for each set of informants. The model fit very well for the BFI and for self-ratings on the Mini-Markers. The three sets of peer ratings on the Mini-Markers did not show adequate model fit. In these three cases, however, the fit of the model could be significantly improved, all $\chi^2_{\text{difference}}(1, N = 487) > 20.90$, all $ps < .001$, by allowing the uniquenesses for Neuroticism and Openness/Intellect to correlate (dotted line in Figure 4). Fit indices for all three models were then similar to those obtained for the models of peer ratings on the BFI. The correla-

tions between the Neuroticism and Openness/Intellect uniquenesses were all significant (rs ranged from .54 to .75, all $ps < .001$). This finding is consistent with our previous findings with the TDA, in which the fit of the higher-order factor model was improved by allowing these two uniquenesses to correlate (DeYoung et al., 2002). One possible reason for this correlation is that in some instruments the positive correlation between Stability and Plasticity may cause the higher-order factor model to overestimate the negative correlation between Neuroticism and Openness/Intellect (thus allowing a positive correlation between their uniquenesses to improve the fit of the model). Another possible reason is that in both the TDA and Mini-Markers, Openness/Intellect and Neuroticism are the only scales that do not have balanced keying (specifically, there are more positively than negatively keyed items). This situation could cause them to covary as a result of acquiescence bias.

Discussion

When modeled as latent variables defined by the ratings of four different informants, the Big Five were significantly intercorrelated. Correlations among the Big Five, therefore, cannot be dismissed as artifacts of the biases of individual raters. Because the Big Five were not completely orthogonal in latent space for either the BFI or the Mini-Markers, the question of higher-order factor structure remains relevant. A hierarchical model with latent Stability and Plasticity variables above the latent Big Five fit the data very well for the BFI. The model did not show the standard Plasticity factor for the Mini-Markers, in that Openness/Intellect did not load significantly on it, but this seems likely to be the result of attenuated correlations due to lower interrater agreement. Despite some attenuation, the pattern of correlations among the latent Big Five for the Mini-Markers was similar to that for the BFI, and a model combining the Mini-Markers and the BFI, as indicators of latent Big Five variables for each informant, demonstrated both metatraits. The higher-order factors thus do not appear to be a method artifact specific to the BFI (not surprisingly, given that

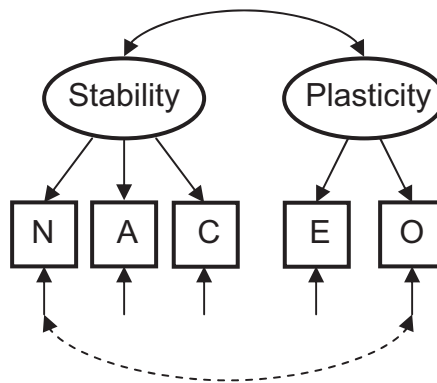


Figure 4. Higher-order factor model for single-informant ratings of the Big Five traits (N = Neuroticism; A = Agreeableness; C = Conscientiousness; E = Extraversion; O = Openness/Intellect). The dotted line represents the correlation between uniquenesses for Neuroticism and Openness/Intellect. Freeing this parameter significantly improved the fit of the model for the three sets of peer ratings on the Mini-Markers.

Table 5
Standardized Parameters and Fit Indices for Higher-Order Factor Models of Single-Informant Ratings (Figure 4)

Measure/informant	Factor loadings					r^a	$\chi^2(4)$	CFI	RMSEA ^b
	N	A	C	E	O				
Big Five Inventory (<i>N</i> = 483)									
Self	-.64**	.53**	.49**	.76**	.33**	.39**	4.68	.99	.019 (.000–.074)
Peer 1	-.76**	.70**	.47**	.46**	.62**	.46**	11.69*	.98	.063 (.023–.107)
Peer 2	-.78**	.67**	.53**	.57**	.60**	.46**	10.89*	.98	.060 (.018–.104)
Peer 3	-.76**	.77**	.58**	.46**	.60**	.49**	9.61*	.99	.054 (.005–.099)
Mini-Markers (<i>N</i> = 487)									
Self	-.49**	.70**	.26**	.58**	.33**	.41**	10.37*	.95	.057 (.014–.101)
Peer 1	-.69**	.79**	.44**	.26**	.57**	.50**	30.02**	.90	.116 (.079–.156)
Peer 2	-.72**	.75**	.37**	.31**	.61**	.45**	33.86**	.90	.124 (.087–.164)
Peer 3	-.67**	.87**	.40**	.12	.67*	.41**	35.39**	.91	.127 (.091–.167)

Note. N = Neuroticism; A = Agreeableness; C = Conscientiousness; E = Extraversion; O = Openness/Intellect. CFI = comparative fit index; RMSEA = root mean square error of approximation.

^a Correlation between stability and plasticity. ^b 90% confidence intervals are presented in parentheses.

* $p < .05$. ** $p < .001$.

almost none of the many previous data sets in which the metatraits were found have used the BFI).

The latent metatraits in the multi-informant models were uncorrelated, in contrast to models fit for single-informant ratings, in which the metatraits were fairly strongly correlated, as in previous studies (DeYoung et al., 2002). Thus, whereas the higher-order factor structure does not appear to be an artifact of the biases of individual raters, the correlation between the metatraits may be artifactual. These findings have implications for research utilizing single-informant ratings: Correlation between the metatraits may suppress associations with other variables, when the metatraits predict in opposite directions. We found in a previous study, for example, that Stability predicted conformity positively, whereas Plasticity predicted it negatively; however, the association with Plasticity did not appear in zero-order correlations and only became evident when controlling statistically for Stability (DeYoung et al., 2002).

Although in the present study the Big Five were correlated and showed the expected higher-order factor structure in latent space, the magnitude of correlations and loadings on the higher-order factors was generally lower than in ratings by single informants. This suggests that individual informants do inflate the correlations among the Big Five, perhaps in part because of a bias toward rating targets' personalities as uniformly desirable or undesirable. However, such a general bias cannot be the only factor leading to the inflation of correlations, because statistical comparisons of the CTCU and CTOM models indicated that the method effects associated with specific informants were not unidimensional for either the BFI or the Mini-Markers.

Exploratory factor analyses of the correlations among uniquenesses were therefore carried out to investigate their factor structure. The uniquenesses represent variance specific to each informant after shared variance has been removed—in other words, the portion of the rating not agreed upon by all four raters. Correlations among the uniquenesses therefore indicate how individual ratings of the Big Five correlate, above and beyond the actual correlations of the traits in latent space. The correlations among

uniquenesses did not show an entirely consistent factor structure. However, for all four ratings of the BFI and for self-ratings of the Mini-Markers, their factor structure was similar to the standard higher-order factors, in which Neuroticism (reversed), Agreeableness, and Conscientiousness mark the first factor and Extraversion and Openness/Intellect mark the second. It appears, therefore, that the biases associated with individual raters generally conform to the same factor structure that is present in the Big Five at the latent level, but that this is more true for the BFI than for the Mini-Markers, perhaps because ratings on the latter are less consistent, as indicated by lower interrater agreement coefficients. This finding suggests that people's expectations about which personality traits should vary together are reasonably accurate (in that their individual biases tend to show the same factor structure as the latent traits) but lead them to attribute more covariation than actually exists, producing inflated correlations and higher-order factor loadings, in single-informant ratings.

One must also consider the possibility that the uniquenesses contain some genuine variance in addition to bias. The latent Big Five variables represent only the variance that is shared among all informants. It is certainly possible that any particular informant may accurately detect some aspect of the target's personality that other informants have overlooked. Such disparities are especially likely when comparing self- and other ratings. Individuals may know things about themselves, through introspection, that others do not. Similarly, access to a more objective view of an individual's behavior may lead others to notice (or report) regularities that the individual does not. This hypothesis is supported by the finding that self- and other ratings yield incremental validity in predicting important criterion variables, such as job performance (Mount, Barrick, & Strauss, 1994). The true magnitudes of the correlations among the Big Five, therefore, probably fall somewhere between those seen in single-informant ratings and those seen in the shared variance of ratings by multiple informants. Correlations in single-informant ratings are presumably higher than they should be, because of various biases, but correlations among latent variables derived from multiple-informant ratings may be lower than they

should be, because of exclusion (from the latent variables) of genuine variance not detected by all informants.

The pattern of correlations among the Big Five might be affected by at least three other factors, which could be examined in future research. First, stronger correlations in latent space might be detected with multiple-informant ratings on an instrument prone to even higher levels of interrater agreement than the BFI, such as the NEO PI-R (Costa & McCrae, 1992a). Second, the breadth of assessment within each Big Five domain might also influence correlations, and here again the NEO PI-R would be useful, as it assesses a wider range of content within each domain than either the BFI or the Mini-Markers. Third, the fact that the vast majority of peer raters in this study felt very positively about their targets could have had some effect on the correlational structure of ratings, and it would be informative to attempt a replication in a more evaluatively heterogeneous sample.

Comparison of Instruments

The average magnitude of the correlations and the number of significant correlations among latent Big Five variables were significantly greater for the BFI than for the Mini-Markers. This seems likely to be due to the greater interrater agreement associated with the BFI and may explain a previous failure to detect significant correlations among the Big Five in a similar multi-informant analysis (Biesanz & West, 2004). Biesanz and West (2004) found no significant correlations among latent Big Five variables in a data set comprising self-, peer, and parent ratings. However, the interrater agreement obtained in that study was lower than that obtained for either instrument in the present study.

Some of these differences in interrater agreement may be due to choice of instrument. Biesanz and West (2004) used a single-adjective rating instrument (the TDA) containing a number of difficult and unfamiliar adjectives (precisely those adjectives that were removed in the construction of the Mini-Markers; Saucier, 1994), which is likely to reduce the consistency of interpretation of items. Of course, other factors related to the participants or the relationships between raters and targets may have contributed to low interrater agreement in their sample. Nonetheless, the present study demonstrated, within one sample, that a single-adjective rating scale (the Mini-Markers) had lower interrater agreement than a scale embedding trait-descriptive adjectives in longer phrases (the BFI) and that interrater agreement was associated with the strength of correlations among the Big Five. At the very least, the results of the current study suggest that one should be attentive to interrater agreement when using multi-informant ratings as indicators of latent variables.

At least two strategies could be used in future research, to strengthen the conclusions of this study regarding the different properties of different instruments. First, as mentioned above, it would be ideal to replicate the current findings with another instrument possessing relatively high interrater agreement, such as the NEO PI-R. Second, examining two samples showing different levels of interrater agreement despite using the same instrument would be of interest. Thus far, three instruments have been used to conduct multi-informant analyses of the Big Five, two in the present study and one by Biesanz and West (2004). Across these analyses, interrater agreement has been perfectly correlated with the average absolute correlation among the Big Five at the latent

level, but each analysis has used a different instrument, thereby confounding the effect of interrater agreement with the effect of instrument. In other words, because neither instrument nor level of agreement among raters has been held constant across analyses, one cannot assert with confidence which of these two factors is responsible for differences in the magnitude of correlations.

Finally, one should consider two additional factors that might contribute to the finding that correlations among the Big Five as assessed by the Mini-Markers were lower than those as assessed by the BFI. First, the Mini-Markers were intentionally designed to produce relatively weak interscale correlations in single-informant ratings (Saucier, 1994). Second, differences in item content between the two instruments might affect Big Five intercorrelations. One salient example is related to the attenuated correlation of Extraversion and Openness/Intellect seen in the Mini-Markers, which precluded the loading of Openness/Intellect on Plasticity. In the Openness/Intellect scale, the Mini-Markers contain more items emphasizing intellectuality than does the BFI, and the BFI includes items related to curiosity and dislike of routine, whereas the Mini-Markers do not. Curiosity and dislike of routine seem likely to be more strongly related to Extraversion than is intellectuality, and their inclusion might make the BFI's Openness/Intellect scale a better indicator of Plasticity than the comparable scale of the Mini-Markers.

The Meaning of the Metatraits

Having provided evidence that correlations among the Big Five are real and appear to possess the higher-order factor structure first reported by Digman (1997), we now return to questions of interpretation and explanation of the higher-order factors. The present findings suggest that although some of the variance in the metatraits in single-informant ratings is a method artifact stemming from the biases of individual raters, enough of it is genuine that the existence of the metatraits must be taken seriously.

At least two additional reasons exist to consider the metatraits important: First, similar higher-order factors have been found in factor analyses combining various measures of normal and abnormal personality traits in conjunction with standard Big Five instruments (Markon et al., 2005). Markon et al.'s analyses indicate that (a) traits considered pathological can be located within the same hierarchy of classification as normal traits and (b) two broad classes of psychopathology (internalizing problems such as depression and anxiety and externalizing problems such as aggression and impulsivity) are associated with low Stability. Second, factors bearing an obvious resemblance to Stability and Plasticity appear in lexical studies when only two factors are extracted (Saucier, 2003; Saucier, Georgiades, Tsoussis, & Goldberg, 2005). These two lexical factors, often labeled Social Propriety and Dynamism, show greater cross-language replicability than do the Big Five (Saucier et al., 2005). A higher-order factoring approach starting with the Big Five thus appears not to be the only method for observing that personality descriptors cluster into two very broad domains. The exact degree to which the Social Propriety and Dynamism factors are similar to the Stability and Plasticity factors is a question for future research. The methods used to discover the two sets of constructs are different enough that one cannot yet judge whether their different labels reflect genuinely different content.

We chose the labels Stability and Plasticity to replace Digman's (1997) "provisional" (p. 1248) labels, α and β , because they seem to be good descriptors of the very general patterns of behavior and experience indicated by the shared variance of Neuroticism (reversed), Agreeableness, and Conscientiousness, on the one hand, and Extraversion and Openness/Intellect, on the other (DeYoung et al., 2002). We have noted elsewhere that this interpretation seems compatible with Digman's (1997) suggestion that α and β might be associated with socialization and personal growth, respectively. Stability seems likely to make a child easier to socialize (and socialization may encourage Stability), whereas Plasticity seems likely (though not inevitably) to lead to personal growth (DeYoung et al., 2002, 2005). Socialization and personal growth, however, seem more like outcomes than predispositions, whereas "Stability" and "Plasticity" suggest more basic tendencies.

We have argued that Stability and Plasticity might be related to two fundamental human concerns (DeYoung et al., 2005): (a) the need to maintain a stable organization of psychosocial function and (b) the need to explore and incorporate novel information into that organization, as the state of the individual changes both internally (developmentally) and externally (environmentally). On this interpretation, some of the variance in the Big Five represents individual differences in emphasis on and competence in meeting these two needs: An absence of Neuroticism reflects emotional stability. Agreeableness reflects the tendency to maintain stability in social relationships (cf. Graziano & Eisenberg, 1997). Conscientiousness appears to reflect motivational stability, the tendency to set goals and work toward them in a reliable and organized manner. Extraversion reflects sensitivity to the possibility of reward (Depue & Collins, 1999; Lucas, Diener, Grob, Suh, & Shao, 2000), producing the tendency to explore the world through action (of course, much of the human world is social, and speech is a form of action). Openness/Intellect reflects the tendency to explore the world perceptually and cognitively (DeYoung et al., 2005). Consistent with our interpretation of Plasticity, both Extraversion and Openness/Intellect are positively related to sensation seeking (Aluja, Garcia, & Garcia, 2003).

The present findings suggest that "Stability" is a good label in part because of its similarity to "emotional stability," the standard label for the negative pole of Neuroticism. In all three multi-informant models that included the metatraits, the loading of Neuroticism on Stability was approximately -1.00 , with weaker loadings for Agreeableness and Conscientiousness. Studies using only single-informant ratings have found loadings for Neuroticism to be lower ($\sim .60$) and similar to loadings for Agreeableness and Conscientiousness (DeYoung et al., 2002; Digman, 1997). The present study reveals that unless the shared variance across informants in this study seriously underestimates the correlation between Agreeableness and Conscientiousness, emotional stability appears to be the primary and dominant component of Stability.⁷ Nonetheless, Stability is conceptually broader than low Neuroticism because it encompasses those aspects of Agreeableness and Conscientiousness that vary with Neuroticism.

The term Plasticity, to denote a broad tendency toward exploration, provides a good complement to Stability, especially with reference to information-processing theory (DeYoung et al., 2002). On the basis of his work with neural network models, Grossberg (1987) used these terms to describe two partially independent subsystems that he argued were necessary for any complex

information-processing system to function well over time in a changing environment: a stability subsystem responsible for maintaining the stability of classification and output and a plasticity subsystem responsible for handling novel information and adjusting categories. The needs met by the functions of these two subsystems seem strongly analogous to the needs described above as the conceptual basis for the traits of Stability and Plasticity.

For any interpretation of the metatraits, an important question is how they might be instantiated biologically. Numerous studies have demonstrated that the Big Five show substantial heritability, with at least 40–50%, and perhaps as much as 80%, of their variance stemming from genetic sources (Bouchard, 1994; Loehlin, 1992; Reimann, Angleitner, & Strelau, 1997). Ample evidence indicates that environmental forces also influence the Big Five over the life span (Roberts, Wood, & Smith, 2005), but environmental forces that affect personality may do so by affecting brain systems, and the question of how traits are instantiated biologically is therefore partially distinct from the question of whether their distal sources are genetic or environmental (DeYoung et al., 2005).

Though nonbiological forces may be partially responsible for trait correlations, patterns of covariance among traits are nonetheless useful as clues to the neurobiological underpinnings of personality. The existence of the metatraits suggests that their constituent Big Five traits may share some aspects of their biological substrates. In previous work (DeYoung et al., 2002), we reviewed evidence supporting the hypotheses that Stability reflects individual differences in the functioning of the serotonergic system, which regulates the stability of emotion and behavior (Spooit, 1992; Zald & Depue, 2001) and that Plasticity reflects individual differences in the functioning of the dopaminergic system, which governs exploratory behavior and cognitive flexibility (Ashby, Isen, & Turken, 1999; Braver & Barch, 2002; Depue & Collins, 1999; Panksepp, 1998; Peterson, Smith, & Carson, 2002). Note that this model is not intended to imply that two neurotransmitters might constitute the entire neurobiological substrate of a trait hierarchy based on the Big Five. (We have presented a model of Openness/Intellect, for example, linking it not only to dopaminergic function but also to the functions of the dorsolateral prefrontal cortex; DeYoung et al., 2005). Undoubtedly, many other neurobiological systems are involved in personality (Zuckerman, 2005). What the model does imply, however, is that individual differences in serotonergic and dopaminergic function are likely to be at least partially responsible for the pattern of correlations among the Big Five. Serotonin and dopamine act very broadly in the brain as neuromodulators, and their effects on personality might therefore be expected to be evident at a level of organization as broad as the metatraits.

That Neuroticism and Agreeableness showed the strongest correlation among the latent Big Five, for both the BFI and the

⁷ The importance of emotional stability within Stability suggests a possible link to Tellegen's higher-order factor Negative Emotionality as does the strong negative correlation between Neuroticism and Agreeableness at the latent level, given the association of Negative Emotionality with aggression (Tellegen & Waller, 1994). Digman (1997) noted this parallel as well. Future studies should empirically compare Stability and Plasticity with higher-order factors from personality models not derived from the lexical tradition (cf. Markon et al., 2005).

Mini-Markers (Table 3), deserves particular attention because molecular genetic evidence indicates that covariance between these two traits, in self-reports, is partially mediated by a specific individual difference in the serotonergic system. Jang et al. (2001) found that the correlation between Neuroticism and Agreeableness was genetically based and that variation in the serotonin transporter gene accounted for 10% of the genetic correlation. This kind of genetic investigation, or investigations utilizing pharmacological manipulations, will be necessary to test our biological model of the metatraits directly. For now, the model remains a plausible hypothesis, synthesizing existing findings and suggesting avenues for further research. Notably, Markon et al.'s (2005) finding of association between Stability and both internalizing and externalizing problems is consistent with our biological model, given that both internalizing and externalizing problems are associated with low levels of serotonin (Spont, 1992). Our model may therefore aid in linking the biological substrates of normal and abnormal variation in personality.

Conclusion

The multi-informant analysis presented here provides evidence that correlations among the Big Five, in two commonly used instruments, are not due to raters' biases. Other instruments seem likely to replicate this pattern, given adequate interrater agreement. One likely exception that should be noted is the orthogonal marker sets developed by Saucier (2002), which yield orthogonal Big Five scores in single-informant ratings. Presumably, these instruments would also yield orthogonal latent traits in a multi-informant analysis like the present one. However, the fact that it is possible to create orthogonal Big Five instruments, through careful item selection, does not necessarily entail that such instruments are desirable. If the Big Five are truly correlated trait domains, then orthogonal marker sets may misrepresent their content. The traditional conception of the Big Five as orthogonal is partly a historical accident resulting from the methods used in their discovery. Had more of the factor analyses that originally validated the Big Five model been performed with oblique rotations that allow correlations among factors, instead of orthogonal rotations that artificially prevent any correlations among factors (at the expense of explaining less variance), we might never have had to debate the reality of correlations among the Big Five. Rather than attempting to eliminate these correlations through orthogonal rotation or techniques of scale construction, one might consider instead whether these correlations, and the higher-order factors they reveal, have substantive meaning.

Our model of Stability and Plasticity offers one interpretation of their meaning and provides a hypothesis regarding their biological sources that may aid in the development of neurobiological theories of personality based on the Big Five. Note that this model does not imply that the metatraits should supplant the Big Five as the most important level of trait organization. In the multi-informant models reported here, neither the correlations among the latent Big Five nor most of the higher-order factor loadings were particularly strong. We have noted elsewhere that Extraversion and Openness/Intellect are probably more different than similar (DeYoung et al., 2005), and the same could be said for Neuroticism, Agreeableness, and Conscientiousness. What is unique to each Big Five trait needs explaining just as much as what is shared. To develop either sort

of explanation, however, one must differentiate what is shared from what is unique, and this can be accomplished only if correlations and higher-order factors are acknowledged and taken into account. The current study suggests strongly that correlations among the Big Five are substantively real and possess a meaningful higher-order structure.

References

- Aluja, A., Garcia, O., & Garcia, L. F. (2003). Relationships among Extraversion, Openness to Experience, and sensation seeking. *Personality and Individual Differences, 35*, 671–680.
- Arbuckle, J. L. (2003). Amos 5.0 (Build 5138). Chicago: SmallWaters.
- Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529–550.
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Blas, L. D., Boies, K., & De Raad, B. (2004). A six-factor structure of personality descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*, 356–366.
- Biesanz, J. C., & West, S. G. (2004). Towards understanding assessments of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality, 72*, 845–876.
- Bouchard, T. J. (1994). Genes, environment, and personality. *Science, 264*, 1700–1701.
- Braver, T. S., & Barch, D. M. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience & Biobehavioral Reviews, 26*, 809–817.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociological Methods and Research, 29*, 468–508.
- Costa, P. T., Jr., & McCrae, R. R. (1992a). Four ways five factors are basic. *Personality and Individual Differences, 13*, 653–665.
- Costa, P. T., Jr., & McCrae, R. R. (1992b). Reply to Eysenck. *Personality and Individual Differences, 13*, 861–865.
- Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences, 22*, 491–569.
- DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences, 33*, 533–552.
- DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of Openness/Intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality, 73*, 825–858.
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*, 1246–1256.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26–42.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26–34.
- Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology, 48*, 82–98.
- Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 767–793). San Diego, CA: Academic Press.
- Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science, 11*, 23–63.
- Jang, K. L., Hu, S., Livesley, W. J., Angleitner, A., Reimann, R., Ando, J., Ono, Y., Vernon, P. A., & Hamer, D. J. (2001). Covariance structure of Neuroticism and Agreeableness: A twin and molecular genetic analysis

- of the role of the serotonin transporter gene. *Journal of Personality and Social Psychology*, 81, 295–304.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford Press.
- Loehlin, J. C. (1992). *Genes and environment in personality development*. Newbury Park, CA: Sage.
- Lucas, R. E., Diener, E., Grob, A., Suh, E. M., & Shao, L. (2000). Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology*, 79, 452–468.
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, 88, 139–157.
- McCrae, R. R., & Costa, P. T. (1995). Positive and negative valence within the five-factor model. *Journal of Research in Personality*, 29, 443–460.
- McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press.
- Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology*, 79, 272–280.
- Mutch, C. (2005). Higher-order factors of the Big Five model of personality: A reanalysis of Digman (1997). *Psychological Reports*, 96, 167–177.
- Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Peterson, J. B., Smith, K. W., & Carson, S. (2002). Openness and Extraversion are associated with reduced latent inhibition: Replication and commentary. *Personality and Individual Differences*, 33, 1137–1147.
- Reimann, R., Angleitner, A., & Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins reared together using the self- and peer report NEO-FFI scales. *Journal of Personality*, 65, 449–476.
- Roberts, B. W., Wood, D., & Smith, J. (2005). Evaluating Five Factor theory and social investment perspectives on personality trait development. *Journal of Research in Personality*, 39, 166–184.
- Saucier, G. (1994). Mini-Markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63, 506–516.
- Saucier, G. (2002). Orthogonal markers of orthogonal factors: The case of the Big Five. *Journal of Research in Personality*, 36, 1–31.
- Saucier, G. (2003). Factor structure of English-language personality type-nouns. *Journal of Personality and Social Psychology*, 85, 695–708.
- Saucier, G., Georgiades, S., Tsaousis, I., & Goldberg, L. R. (2005). The factor structure of Greek personality adjectives. *Journal of Personality and Social Psychology*, 88, 856–875.
- Saucier, G., & Goldberg, L. R. (2001). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality*, 69, 847–879.
- Spoont, M. R. (1992). Modulatory role of serotonin in neural information processing: Implications for human psychopathology. *Psychological Bulletin*, 112, 330–350.
- Tellegen, A., & Waller, N. (1994). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In S. R. Briggs & J. M. Cheek (Eds.), *Personality measures: Development and evaluation* (Vol. 1, pp. 133–161). Greenwich, CT: JAI Press.
- Yik, M. S. M., & Russell, J. A. (2001). Predicting the big two of affect from the Big Five of personality. *Journal of Research in Personality*, 35, 247–277.
- Zald, D. H., & Depue, R. A. (2001). Serotonergic functioning correlates with positive and negative affect in psychiatrically healthy males. *Personality and Individual Differences*, 30, 71–86.
- Zuckerman, M. (2005). *Psychobiology of personality, second edition*. New York: Cambridge University Press.

Received June 14, 2005

Revision received March 21, 2006

Accepted June 16, 2006 ■