A Phenotypic Null Hypothesis for the Genetics of Personality

Eric Turkheimer,[1] Erik Pettersson,[2] and Erin E. Horn[1]

[1]Department of Psychology, University of Virginia, Charlottesville, Virginia 22904; email: ent3c@virginia.edu

[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE-171 77 Stockholm, Sweden

Annu. Rev. Psychol. 2014. 65:515–40

First published online as a Review in Advance on September 18, 2013

The *Annual Review of Psychology* is online at http://psych.annualreviews.org

This article's doi: 10.1146/annurev-psych-113011-143752

Keywords

behavior genetics, twins, genomics

Abstract

We review the genetically informed literature on the genetics of personality. Over the past century, quantitative genetic studies, using identical and fraternal twins, have demonstrated that differences in human personality are substantially heritable. We focus on more contemporary questions to which that basic observation has led. We examine whether differences in the heritability of personality are replicable across different traits, samples, and studies; how the heritability of personality relates to its reliability; and how behavior genetics can be employed in studies of validity, and we discuss the stability of personality in genetic and environmental variance. The appropriate null hypothesis in behavior genetics is not that genetic or environmental influence on personality is zero. Instead, we offer a phenotypic null hypothesis, which states that genetic variance is not an independent mechanism of individual differences in personality but rather a reflection of processes that are best conceptualized at the phenotypic level.

Contents

INTRODUCTION

Personality and behavior genetics have a special relationship. The scientific origin of both fields is in the nineteenth century, and they came of age at the same time, after World War II, as human personality was distinguished from cognitive ability on the one hand and psychopathology on the other and as behavior genetics embarked on its modern empirical programs of experimental studies of model organisms and quantitative genetic studies of humans. Both personality psychology and behavior genetics were spurred by the development of modern factor analysis and the computational power that supported it.

Another reason for this special relationship is even more important. The nineteenth-century roots of behavior genetics involved the classical questions of nature and nurture formulated by Francis Galton, questions that are still important to this field. However, for personality, as opposed to other phenotypes such as intelligence and psychopathology, the so-called nature-nurture debate was never an issue. For thousands of years, animal breeders had been selecting domesticated livestock for behavioral traits; any farm owner, never mind any dog owner, knew perfectly well that behavioral traits with strong analogs to human personality could be transmitted genetically in lower animals, even prior to any scientific knowledge about what "genetic" transmission entailed.

The earliest research known as behavior genetics involved transmission and breeding of temperamental traits in dogs. The first article about behavior genetics in this journal (Fuller 1960) extensively covered the genetics of temperament in *Drosophila*, mice, and dogs, with scarcely any consideration of nature and nurture or genes and environment; in the experimental studies of breeding at that time, the unity of nature and nurture was taken for granted. Had the behavior genetics of personality remained focused on experimental studies of temperament in mice and dogs, the field's history would not be nearly as fraught as we find it today. Investigators inevitably decided to extend the incontrovertible research on the genetics of personality in lower animals to the analogs of those temperamental traits in humans. Although traits such as aggression and activity level translate fairly transparently from dogs to humans, the breeding and cross-fostering studies that had been employed to study them do not, so investigators had to turn to other methods, namely the twin and adoption studies that eventually came to be the hallmark of modern behavior genetics.

Phenotype: the observable characteristics of an organism, as opposed to their genetic or environmental origins

Heritability: the proportion of variance in phenotype that is associated with variation in genotype

BASICS OF BEHAVIOR GENETICS

The goal of this article is to go beyond the assertions and denials of heritability that have traditionally characterized the genetics of behavior. A very brief review is necessary, however, to introduce

some terms and abbreviations that are used in the remainder of the article. In the classical twin model, phenotypic variances and covariances of pairs of identical or fraternal twins are partitioned into three components: the additive effects of multiple genes (A), of which 100% are shared in identical twins and 50% in fraternal twins; the shared environmental effects that make siblings raised together in the same family similar (C); and the remainder (E), sometimes termed the nonshared or unique environment, which comprises everything that makes twins raised together different, including measurement error. Some elaborations of this basic design are introduced below.

The assumptions—both statistical and biological—of the classical twin model have been hotly contended for as long as twin studies have existed, and disagreement about them has not abated (Joseph 2004, Charney 2012). We do not use our limited space debating these issues, for several reasons. They have, of course, been debated many times already. In addition, objections to the assumptions of twin studies are most relevant when the goal of the studies is to compute the heritability of one trait or another, and our explicit goal is to avoid doing so. We have made the case elsewhere (Turkheimer 1998, 2000; Turkheimer & Harden 2013) that the numerical values of heritability coefficients do not matter very much anyway, other than by differing from zero or one. Moreover, some recent DNA-based statistical methods that do not require twins or any assumptions about them have reached conclusions very similar to those from the classical twin studies (Turkheimer 2011, Yang et al. 2011).

With that in mind, we now turn to the question of whether differences in human personality are heritable. We can be mercifully brief: yes. Every review of the genetics of personality, from the early reports from Cattell (1981) and Eysenck (1990) to modern summaries by Plomin & Caspi (1990), Bouchard & Loehlin (2001), and Krueger & Johnson (2008), has concluded that identical twins are more similar for personality traits than are fraternal twins and that the personalities of adopted children are more similar to the personalities of their biological parents than to those of their adoptive parents. Personality is not alone in this regard; indeed, Turkheimer (2000) has long argued that all human traits are heritable, referring to the universality of heritability as the First Law of Behavior Genetics.

The other two laws of behavior genetics pertain to the two environmental components of the classical model: the shared and nonshared environment, and there are some basic results regarding them that should be discussed before proceeding to other questions. The Second Law of Behavior Genetics, which states that the shared environmental component of human individual differences is small, is usually true for most traits, but the situation is somewhat starker for personality. It is remarkable, in surveying the genetically informed personality literature in a very wide context, how completely absent the shared environment is. In fact, it is often the case that identical twins are more than twice as similar as fraternal twins, a violation of the classical twin model that, if uncorrected, produces negative estimates for shared environmental variance. In this review, the near-unanimous absence of shared environmental effects provides a useful simplifying assumption that allows us to focus on genetic effects (which we sometimes refer to simply as A) and nonshared environmental ones (E) (see sidebar Why Are There No Shared Environmental Effects on Personality?).

The Third Law of Behavior Genetics states that even identical twins raised in the same home are not perfectly correlated for anything, especially behavior and certainly not personality. Uncorrelated variance between members of an identical twin pair is known as the unique or nonshared environment, and although we use the latter term here it is misleading in many ways (Turkheimer & Waldron 2000). We prefer to consider the nonshared environment in more concrete terms, as the phenotypic variance within identical twin pairs raised together, especially as an alternative to thinking of it as some unspecified set of environmental agents that cause members of identical twin pairs to differ from each other. We apply this distinction to the analysis of validity studies in the remainder of this review, and hopefully its utility will become apparent.

WHY ARE THERE NO SHARED ENVIRONMENTAL EFFECTS ON PERSONALITY?

One possibility is that complex genetic interactions [epistasis, or what Lykken (1982) has referred to more broadly as emergenesis] produce configural effects that increase the similarity of identical twin pairs compared with all other types of relationships. Loehlin and colleagues (2003) have shown that analyses including half-siblings demonstrate surplus similarity in identical twins relative to other relationship types. One must also consider the possibility, however, that families simply do not contribute much common systematic variance to the personalities of children raised together. In the domain of cognitive ability, hypotheses about the absence of family effects are fraught with controversy, for good reasons. Intelligence is a directional trait; in general it is always good to have more of it, and parents invest extraordinary resources in the cognitive abilities of their children. One of the most important social institutions in modern civilization—the educational system—is dedicated to increasing cognitive ability in children, and varies mostly at the level of families (i.e., children raised in the same family are usually exposed to the same schools). Personality traits, in contrast, are bidirectional, with positive and negative traits at both ends, and there is nothing analogous to the educational system dedicated to changing them.

VARIABILITY OF HERITABILITY

Are some personality traits more heritable than others? This would seem to be a foundational issue of behavior genetics as it has traditionally been formulated. If the goal of behavior genetics is to answer nature-nurture questions, then one would expect the answers to the questions to differ, trait by trait. Unfortunately, this particular issue suffers from widely acknowledged but frequently ignored limitations inherent in the concept of heritability itself. We recently discussed this issue at length (Turkheimer & Harden 2013) and do so only briefly here. Reviews of the heritability concept always include the caveat that a heritability coefficient applies only to the population in which it was computed, but the most important implications of this limitation are not generally acknowledged.

A heritability coefficient represents the proportion of phenotypic variability that is associated with variability in genotype. As such, it is an effect size, a variance ratio, an R^2 coefficient; and like any variance ratio it is sensitive to characteristics of the population in ways that means are not. In particular, variance ratios depend crucially on the variability of both the predictor and the outcome. For example, the question, "How much of the variance in college performance is explained by differences in SAT scores?" has no meaningful answer, other than, "It depends on the variability of SAT scores and other factors at the institutions where the study is conducted." The dependence of standardized correlation coefficients on their variability is a direct consequence of their presumed advantage, which is that they are unit free. Correlations between x and y are not expressed in units of x and units of y; they are expressed in standard deviations of x and standard deviations of y, and the value of the correlation changes as those standard deviations change. This consideration was the basis of Tukey's (1954) famous opposition to correlation coefficients, as summarized in Turkheimer & Harden (2013).

Notwithstanding these concerns, there is a considerable literature on what is usually termed the differential heritability of personality traits. This literature was initiated by a review by Thompson & Wilde (1973). Thompson was a founder and later president of the Behavior Genetics Association. After reviewing the experimental and animal literature in a manner typical for the time, these authors turned to twin studies, and then to twin studies of personality. In reviewing the extant literature, they noted a number of attempts to "replicate" heritabilities across the genders or ages of twins, and to their apparent surprise the results were uniformly unsuccessful. Rank-order

Genotype: a collective term for the genetic characteristics of an organism

correlations among heritabilities across gender and age ranged from 0.06 to 0.29, did not reach statistical significance, and were as likely to be negative as positive. Dismayed by these results, these authors reached generally negative conclusions about the genetics of personality and the prospects for twin studies in general. The review appears to have spurred the twin research community to take a serious look at the problem, largely in the form of a 30-year research program conducted by Loehlin. Beginning with the classic book *Heredity, Environment, and Personality*, Loehlin & Nichols (1976) conducted an exhaustive analysis of California Psychological Inventory (CPI) scores in a sample of 850 pairs of twins who had taken the National Merit Scholarship Qualification Test (NMSQT).

Loehlin and Nichols's decisive answer was that the relative magnitudes of heritabilities did not replicate. The authors divided the sample by gender, divided the male and female samples into two random subsamples, computed the difference between the identical and fraternal twin correlations in each of the four subsamples, and compared the rank differences from lowest to highest. The pairwise Spearman rank correlations between the subsamples ranged from −0.22 to +0.30; none of them were significantly different from zero. To test whether this result might have occurred because of inadequacies in the CPI scales, these authors constructed their own by using a cluster analysis to create 70 small groupings of three or four items. The results for these scales were no different. They concluded, "In short, for personality and interests, as for abilities, the existing twin literature appears to agree with our own finding that while identical-twin pairs tend to be more similar than fraternal-twin pairs. . . . [t]he difficulty is in showing that trait X is more heritable that trait Y" (Loehlin & Nichols 1976, p. 46).

The subsequent literature did little to change Loehlin and Nichols's conclusion. Carey et al. (1978) reexamined Loehlin and Nichols's results in combination with other samples and demonstrated that monozygotic (MZ) twin correlations were more stable than dizygotic (DZ) ones but that both displayed some detectable stability across samples and that extraversion scales appeared to have slightly higher heritabilities than others. However, as Loehlin (1978) pointed out, the consistencies of the heritabilities were still zero in Carey et al.'s data, just as they were for Loehlin & Nichols (1976).

Loehlin (1982) then returned to the problem, armed with two new tools: a much larger sample (13,000 Swedish twin pairs with measures of extraversion and instability) and structural equation modeling, the application of which to twin studies Loehlin pioneered. Analyses of the Swedish sample suggested that genetic and shared environmental parameters were not equal across the male and female samples or across the three birth cohorts included in the full sample. This apparent success led to another problem, one that continues to be important below: With sufficiently large samples, null hypotheses are always wrong (Meehl 1967). The goal of conducting hypothesis tests in individual studies is not simply to reject or fail to reject hypotheses one at a time but rather, through replication, to build individual hypotheses into cumulative theories that explain the phenomena of interest; the latter goal is much more difficult to achieve than the former. Statistical significance notwithstanding, what is one to make of the finding that the heritability of extraversion in males changes from 0.50 in the earliest-born cohort to 0.36 in the second cohort to 0.66 in the third? And why is the heritability of instability higher for females than for males in two cohorts, but equal in the third?

Loehlin's (1982) other finding in the Swedish sample was that the heritabilities of extraversion and instability could not be differentiated, and that led him to formulate a new hypothesis in the NMSQT sample. Extraversion and neuroticism are the two largest factors in the personality domain, and if their heritabilities are equal, then their relative dominance in the factor matrix could mask differences on less important traits. Returning to the NMSQT data, Loehlin created item clusters, factor-analyzed them, and rotated the first two factors to extraversion and neuroticism.

Five Factor Model (FFM): the predominant model for individual differences in human personality; the five factors are openness, conscientiousness, extraversion, agreeableness, and neuroticism (OCEAN)

The remaining factors (stereotyped masculinity, intolerance of ambiguity, persistence, cynical attitudes, and intellectual interests) showed significant (although, once again, not especially systematic) gender differences and significant differences among the traits. In particular, intolerance of ambiguity and stereotyped masculinity showed lower heritabilities than did other traits, as well as stronger shared family influence.

Several years later, Loehlin (1985) returned to the problem again, this time combining the NMSQT sample with the Veterans Administration Twin Sample (Horn et al. 1976) and adoption data from the Texas Adoption Project. Loehlin analyzed whether identical biometric parameters could be fit to all 18 of the CPI subscales and found fairly decisively that they could not. Once again, it proved difficult to theorize about what the nature of those differences might be. Loehlin (1985, p. 217) concluded, "One could pursue matters further by continuing to fit models on an ad hoc scale-by-scale basis, but in doing so one would presumably be running an increasing risk of merely fitting to idiosyncrasies in the data, so it is perhaps prudent to stop at this point." A further analysis of high- and low-heritability items from the NMSQT showed no consistency with a similar analysis conducted by Horn et al. (1976).

Finally, 30 years after he began this research, Loehlin (2012) revisited the problem in a sample of 2,600 Australian twin pairs, using his original methodology of comparing MZ and DZ twin correlations across male and female pairs divided into two random subsamples. As before, Loehlin cluster-analyzed the items, deriving 11 clusters, including 1 extraversion cluster, 2 neuroticism-like clusters, and various narrower clusters. This time, he found substantial consistency in MZ-DZ differences across the four groups. The biometric results did not vary much across scales; shared environmental terms were zero throughout, and the genetic terms ranged from 0.48 to 0.20. Loehlin noted that the greatest differences in heritability were observed, as before, for the traits that load most highly on broad factors of extraversion and neuroticism, which did not differ from each other.

What can we make of these attempts to find differential heritability of personality traits? The most reliable traits—the ones that account for the most variance in the covariance matrix of personality responses—are the traits for which heritability is least variable. Less reliable traits that account for less variance in the personality matrix are more variable, and thus more likely to differ from each other, but rarely systematically. This pattern of results is typical for all of behavioral genomics. One can identify broad dimensions of behavior; quantify their relation to a broad spectrum of genes; and obtain consistent, replicable results that fail to differentiate among behaviors and become uninteresting once they are established. Under most circumstances, both extraversion and neuroticism are heritable at approximately 0.4, and there is little more to be said. Alternatively, one can focus on narrow domains of behavior or (as in the section titled Genomics of Personality below) the relations of behavior to specific as opposed to agglomerated genetic variance, and obtain results that appear to differentiate among traits or genes but fail to replicate in the next study.

HERITABILITY AND RELIABILITY

Personality assessment is inherently hierarchical. In the Five Factor Model (FFM), each major trait is subdivided into facets. In most classical research on the structure of personality, the facets, and often even the factors themselves, were measured by simply summing responses to individual items. Correspondence between items and scales was determined by (*a*) classical psychometric theory and coefficient alpha, (*b*) a priori groupings of items known as testlets, or (*c*) cluster-analytic methods such as those used by Loehlin. With the advent of item response theory and categorical

factor analytic models on the one hand, and increased computational power on the other, however, there is no reason for the factor-analytic process not to begin with the items themselves, organized hierarchically into facets that are in turn organized hierarchically into traits. In the other direction, the FFM traits are often analyzed into two broader factors, alpha and beta (Digman 1997, DeYoung 2006), and beyond that even into a single general factor of personality (GFP) (Rushton et al. 2008; however, see Pettersson et al. 2012 for a skeptical view of the substantive basis of the GFP).

We characterize this process as one of reliability because the core question, about how personality items group together into traits, is essentially psychometric. Reliability refers to the tendency for multiple measures of a single personality trait to covary. In classical twin models, just as one can partition the variance of a single trait into biometric components, one can also decompose the covariances among multiple traits, the common factors that those covariances define, and the residual variances (error variance, in classical psychometrics; uniqueness, in factor-analytic terminology) of the items after the common variance has been accounted for. The reliability coefficients of classical psychometric theory involve the ratio of the variance of the common factor to the full variances of the items. The behavior genetic question is about the biometric composition of the common factor and the residuals.

Loehlin et al. (1998) investigated common and unique variance in three different measures of the FFM. In this case, common variance in each facet represents multimethod variance among three methods employed in the NMSQT study: self-rating scales, personality inventory items, and adjective checklists. As expected, the common variance in the FFM traits consisted of A and E. The variance unique to the methods had significant but substantially lower heritabilities and was generally more unstable; even the shared environmental term occasionally appeared. Kandler et al. (2010) reported similar results for common and unique variance among self- and peer ratings of personality.

Jang et al. (1998) analyzed common and unique variance among the FFM facets composing the FFM traits and showed that the five main traits of the FFM were heritable at approximately 0.5, whereas the heritabilities of the unique variances of the facets were once again lower but significant. When the components of unique variances were corrected for unreliability, they were indistinguishable from the traits. Jang et al. (2002) administered the NEO PI-R (Neuroticism-Extroversion-Openness Personality Inventory, Revised) and analyzed common and unique variance at the factor and facet levels. For the set of six facets belonging to the same factor, they fit two common A and two common E factors and also partitioned the unique variance of each facet into A and E. Results showed that all traits are around 50% heritable; approximately half the variability in facets is shared with the common factors; shared and nonshared (A and E) variance exists at all levels of the factor hierarchy; more of the common variance is shared in comparison to the unique variance (the heritability of the common variance is higher); and conversely, more of the shared variance is common. In the other direction, by examining higher-level common factors of the FFM, Jang et al. (2006) showed that the higher-order factors of the FFM, alpha and beta, fit the same pattern: Heritabilities are somewhat higher at the facet level than at the trait level but are still substantially lower than unity.

In summary, biometric models of the psychometric structure of personality show that there is heritable variance all the way down to the item level and nonshared environmental variance all the way up to the most general level. The proportion of genetic variance increases as one moves up the psychometric hierarchy, as more and more error of measurement is eliminated, but when reliability is accounted for, the proportion of heritable variance does not seem to vary substantially by level of analysis.

VALIDITY

We have written extensively about the role behavior genetics can play in the assessment of validity (Turkheimer & Harden 2013). The central problem in assessing the validity of personality measures in humans is the evaluation of causal hypotheses, as well as the limitations placed on causal inference by the impossibility of random assignment to experimental conditions. Suppose one hypothesizes that extraversion is a risk factor for illicit drug use, and observes a correlation between measures of the two traits in the general population. Obviously, one cannot conclude from such data that extraversion causes drug use, and most of the experimental tools that might be available with nonhuman participants—everything from cross-fostering studies to random assignment, to drug exposure, to genetic knockouts—are not available to a researcher concerned with humans.

There are two main threats to the validity of causal inferences based on phenotypic associations. The first is direction of causation, the possibility that drug use causes extraversion instead of the other way around. Although there are behavior genetic models that can discriminate direction of causation, at least in theory (Heath et al. 1993), they have proven difficult to apply in practice. Other quasi-experimental methods, particularly longitudinal designs (which, of course, can be combined with genetically informative data), are more practical for concerns about direction of causation. In the remainder of this section, we assume that it is reasonable to presume that the direction of causation flows from personality to some outcome in another domain.

The other kind of threat to causal inferences about phenotypic associations between personality variables and other outcomes involves third-variable confounds. Returning to the example of extraversion and drug use, the genetic background that contributes to extraversion may also contribute to propensity for drug use. If the phenotypic association between extraversion and drug use is mediated genetically, then there is no reason for the more extraverted member of an MZ pair to be more prone to drug use than her introverted cotwin. If, however, extraversion actually causes drug use there is no reason the process would not occur just as clearly within twin pairs as between them. It is crucial to understand that, in this context, genetic correlations between drug use and extraversion are an alternative to a causal hypothesis.

We reach two conclusions from this analysis. First, the causal relationships of interest to psychologists are almost always phenotypic in nature. If extraversion causes drug use, it matters little how the two phenotypes may be divided into biometric variance components; our hypothesis is that phenotypic extraversion causes phenotypic drug use. Second, the nonshared environment has a special role to play in the assessment of causal hypotheses within genetically informed designs. It is useful to consider the nonshared environment in concrete terms, as the difference in phenotype between a pair of identical twins reared in the same family. If, within pairs of identical twins, the twin who is more extraverted is also the twin more likely to use drugs, then the association cannot be mediated by genes, because the twins are genetically identical; it cannot be mediated by a family variable such as neighborhood, because the twins were raised together.

There are many ways to analyze bivariate family designs in which a personality variable is evaluated as a possible cause of an outcome (Turkheimer & Harden 2013). The most straightforward (**Figure 1**) is the so-called bivariate Cholesky decomposition, which corresponds to a biometric regression model in which the phenotypic regression between an outcome and a predictor is decomposed into separate regressions in the ACE domains. We (Turkheimer & Harden 2013) have demonstrated that when the biometric components of the predictor are appropriately (i.e., not) standardized, and when the predictor causes the outcome and is not confounded by uncontrolled A and C processes, then the three regression coefficients, b_A, b_C, and b_E, are equal to one another and to the hypothetical unstandardized structural regression coefficient, b_P (**Figure 1**), expressed as

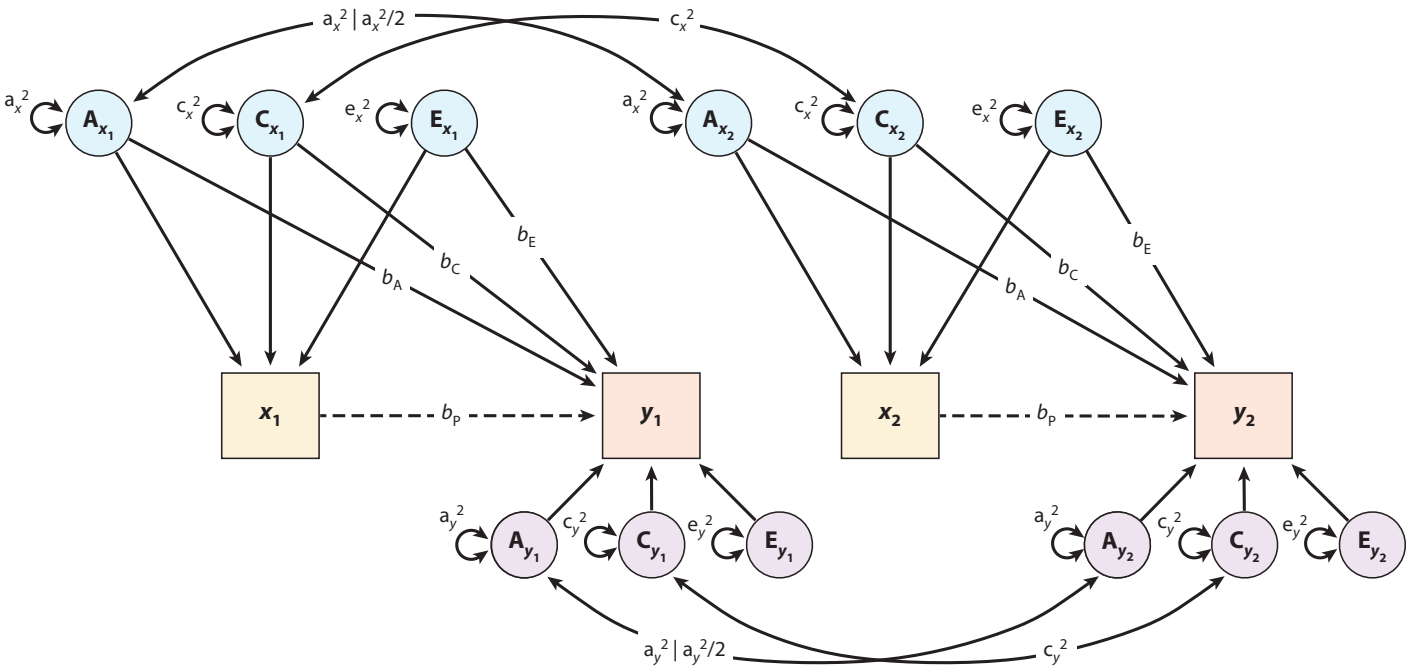

Figure 1

Unstandardized bivariate Cholesky model, representing a genetically informed regression of an outcome on a predictor. The three regressions, b_A, b_C, and b_E, estimate the phenotypic quasi-causal regression b_P, plus genetic and shared environmental confounds. Abbreviations: A, genetic effects; C, shared environmental effects; E, nonshared environmental effects.

phenotypic units of *y* per phenotypic unit of *x*. If there are genetic and shared environmental confounds, and if we assume that the nonshared environmental effect is unconfounded (the crucial assumption of the model), then the nonshared environmental regression continues to estimate b_P; b_A and b_C are equal to b_P plus the magnitude of the genetic and environmental confounds, respectively.

The genetically informed literature on validity in personality is vast and unfocused, encompassing everything with which personality might plausibly be related. We focus on the key issue of the relationship between personality and psychopathology and on the smaller set of studies that report results in three substantive areas in a form similar enough to our unstandardized genetically informed bivariate regression model to allow us to compute the relevant parameters. Klump et al. (2002) analyzed relations between Multidimensional Personality Questionnaire (MPQ) scores [negative emotionality (NE), positive emotionality (PE), and constraint] and disordered eating behaviors [Minnesota Eating Disorders Inventory (M-EDI)] in a sample of twins from the Minnesota Twin Family Study. They reported results in the most commonly used format, a standardized bivariate regression model, in which the regression coefficients have been standardized by fixing the variances of the latent biometric variances of the predictors (in this case, the MPQ scores) to unity. We prefer unstandardized models because the standardized models confound the magnitude of the quasi-causal unstandardized coefficient with the magnitudes of the ACE variances in the predictor (Turkheimer & Harden 2013), but Klump et al. (2002, tables 2 and 3) report the twin correlations in sufficient detail to estimate the unstandardized regressions from the values they report. We describe the results for relations between two higher-order factors of the MPQ (NE and PE; results were not reported for constraint, because the phenotypic relation with M-EDI was not significant) and the total score of the M-EDI.

For NE, a regression of the M-EDI on NE without genetically informative controls shows that a unit of change in NE is associated with 0.28 units of change in the MDI. When the genetic regression is included in the model, however, we observe that much of the observed association is attributable to shared genetic background: The unstandardized genetic regression coefficient is 0.79 units of MDI per unit of NE, and the E regression, estimating the phenotypic effect, is 0.08 units, which does remain significant at $p < 0.05$. For PE, the A regression was equal to −0.24 units of M-EDI per unity of PE, which is highly significant and in the same direction as the phenotypic effect; the E regression was actually in the opposite direction (0.09) and, in our reanalysis based on summary statistics, significantly so ($p = 0.03$). These results suggest that the phenotypic results for the relation between PE and eating disorders are potentially misleading. Twin pairs who, on average, have higher PE scores have lower M-EDI scores; but within twin pairs, the twin with the higher PE score has, if anything, higher M-EDI scores.

Klump et al. (2002, p. 387) concluded, "In general, results suggested that common genetic factors contribute more to relationships between personality and disordered eating attitudes and behaviors than common nonshared environmental factors." This statement is true as far as it goes, but it misses an important point: The "nonshared environmental factors" the authors are discounting are not unknown and unmeasured environmental events, but rather the phenotypic causal effects of personality itself. If differences in PE cause differences in eating, one would expect to observe the effect both between and within pairs. Having observed that PE is associated with eating behavior only between pairs, the most useful conclusion is that the observed association is probably not phenotypically causal but is, instead, the result of a shared genetic background (in the absence of shared environmental variance) between PE and eating.

This pattern of findings is broadly characteristic of the genetically informed literature examining the validity of personality as a predictor of psychopathology. One of the better-established relationships is between neuroticism and depression. Fanous et al. (2002) administered the short form of the Eysenck Personality Questionnaire (EPQ) to obtain neuroticism scores and a structured interview for lifetime depression according to DSM-III criteria for a large sample of male and female twins from the Virginia Twin Registry. These authors report parameters from standardized Cholesky decompositions; once again, we use the reported twin correlations to compute the unstandardized parameters. The phenotypic effect of neuroticism on depression was 0.50 in males and 0.33 in females; both had $p < 0.05$. Once the genetic relationship between the two was accounted for, however, the effects were reduced to 0.30 in males and 0.20 in females; again, both had $p < 0.05$.

Kendler et al. (2006) conducted a similar study of neuroticism and depression in a large sample of Swedish twins. The unstandardized phenotypic regression of depression on neuroticism (recomputed from their published results) was 0.39 in females and 0.40 in males. Once the common genetic background of neuroticism and depression was controlled, the corresponding regression coefficients were 0.13 in males and 0.06 in females.

STRUCTURE

Since the beginning of factor-analytic personality research, investigators have endeavored to combine psychometric methodologies with biometric analyses. The simplest way to combine them is to factor-analyze personality items to derive a latent personality structure (e.g., by using the FFM) and then decompose the resulting latent variables into their biometric components. This methodology has led to the usual conclusions, as we review above in the section titled Heritability and Reliability: Personality factors are more or less equally heritable, shared effects of families are hard to find, and the nonshared environment accounts for a substantial proportion of variance

(although a somewhat lower one at the latent level because unreliability of measurement has been eliminated).

A more interesting method of combining psychometric and biometric analyses of personality responses is to reverse the order of the analyses: One may use a twin design to obtain a biometric partitioning of variation and covariation among personality items in genetic and environmental covariance matrices, and then model the psychometric structure separately in the three domains. The empirical investigation of this possibility once again starts with Loehlin, the NMSQT sample, and the CPI. Loehlin (1987) created 31 item clusters in the same way he had when investigating differential heritability. He then employed standard formulas to compute twin correlations for each cluster, used these to compute the ACE correlation matrices for the 31 clusters, and submitted them separately to exploratory factor analysis. In the A and E domains, Loehlin's (1987) results seemed to accord fairly closely with portions of the (at that time not fully developed) FFM. The genetic factors included versions of extraversion, neuroticism, conscientiousness, and openness. The E factors were less well defined but seemed to include a neuroticism factor and (more weakly) extraversion and conscientiousness factors. As usual, there was not a great deal of C variance in these data, but Loehlin managed to extract two factors, one of which was a somewhat artifactual gender factor that arose because only same-sex pairs were analyzed. As a result, any variation due to gender varied only between pairs, equally, for MZ and DZ twins, so gender effects were counted as C.

Loehlin then replicated these analyses in the Veterans Administration Twin Sample (Horn et al. 1976) and the Texas Adoption Project, rotating the factors from the replication samples to maximum congruence to the NMSQT. The FFM-type factors in the A and E domains replicated fairly well; the C factors, especially the gender factor, did not, because the Veteran twins included opposite-sex DZ pairs. Finally, the A factors in the replication samples could be rotated to at least modest concordance with the A factor from the NMSQT. The conclusion, which is typical of the studies that followed, is quite surprising: Personality structure in the ACE domains seems to be mostly a reflection of phenotypic personality structure, which is to say that some version of the FFM fits fairly well in all three.

Carey & Dilalla (1994) conducted a similar analysis on the same data and reached the same conclusion: "[T]he loadings from the three [A, C and E] matrices are remarkably similar and, moreover, parallel the first three phenotypic factors of the CPI." Heath et al. (1994, p. 770) concluded, "The patterns of loadings . . . were remarkably consistent for Genetic and Environmental Factors 1–4 and quite consistent with what would have been predicted from the observed phenotypic correlations." Livesley et al. (1998), analyzing personality disorders using the Dimensional Assessment Personality Disorder Basic Questionnaire, found concordance coefficients greater than 0.95 for pairs of A, E, and phenotypic factors. Krueger (2000) analyzed 11 scales from the MPQ in a sample of 2,490 pairs of twins from the Minnesota Twin Family Study. He conducted principal components analysis on the phenotypic, A, and E correlation matrices; obtained scores for the sets of principal components; and compared them. All the correlations between phenotypic and genetic components were greater than 0.95, and all the correlations between phenotypic and nonshared environmental components were at least 0.87.

McCrae et al. (2001) collected NEO-PI scores from 1,150 German and Canadian twins. They analyzed biometric structure by using the original Loehlin method and found the usual result of equality across the A, C, and E loadings, which they referred to as the puzzle of parallel structure. Jang et al. (2006) showed what appears to be very high factor concordance between the higher-order structure of the phenotypic, A, and E matrices in the Canadian and German twin samples, plus an additional twin sample from Japan. Yamagata et al. (2006) presented similar results at the facet level; all the correspondence coefficients between phenotypic and A matrices and between phenotypic and E matrices were greater than 0.95. Yamagata et al. (2006, p. 994)

concluded, "These results suggest that . . . the phenotypic five-factor structure is reflective of not only genetic structure, but also environmental structure." However, it may be simpler to conclude that it is the other way around: The genetic and environmental structures of personality are a reflection of phenotypic structure, with little evidence to support differences between genetic and environmental structure or etiology.

Loehlin (2011) returned to his cross-validated cluster analytic methods using data from the Australian Twin Data. Three cross-validated A clusters (emotionality, confidence, reserve) emerged, all of which were repeated in the five cross-validated E clusters. Of the three additional E clusters that cross-validated, one (unscrupulousness) was very similar to an A cluster that just missed the cross-validation standard. Two additional E clusters (with only three items each) did not appear in the A clusters. Loehlin & Martin (2013) used factor analysis of A and E matrices to extract broad factors from seven scales from a combination of the EPQ and the MPQ. The factors were rotated to a general factor (cf. Pettersson et al. 2012) and two supplementary factors that were identified as social conformity and other-dependence. These authors (Loehlin & Martin 2013, p. 761) concluded that "the structure of personality is inherent in the evolved phenotype, and is not the immediate consequence of either genetic or environmental organizing factors."

STABILITY

Lifetime developmental trends have attracted considerable interest in recent nongenetic studies of personality (Caspi et al. 2005), and various aspects of stability and change in the genetics of personality have been examined for some time. There are two main phenotypic issues in the genetics of personality that are usually characterized as changes in the mean and in the rank order of personality traits. Questions about changes in the mean of a trait involve trends in the population mean as a function of age; behavior genetics, with its emphasis on individual differences, has little to offer on this question other than the unsurprising conclusion that variation in slopes of change across age has a genetic component. Rank-order changes refer to analyses of stability, that is, whether individuals who score relatively high on a trait at one age continue to do so later in life. Caspi et al. (2005, p. 466) summarized the phenotypic literature on the stability of personality as follows:

> Test-retest correlations over time (*a*) are moderate in magnitude, even from childhood to early adulthood. Furthermore, rank-order stability (*b*) increases with age. Test-retest correlations (unadjusted for measurement error) increased from 0.41 in childhood to 0.55 at age 30, and then reached a plateau around 0.70 between ages 50 and 70. Rank-order stability (*c*) decreases as the time interval between observations increases, and does not vary markedly (*d*) across the Big Five traits nor (*e*) according to assessment method (i.e., self-reports, observer ratings, and projective tests), or (*f*) by gender.

There are many ways to model longitudinal stability in genetically informative models. We prefer state-trait-error (STE) (Kenny & Zautra 1995) models, which allow partitioning of longitudinal effects into stable effects that are constant over time intervals, simplex effects that have a structured relation with time, and random effects that are uncorrelated in time. **Figure 2** shows a schematic path diagram of an STE model for genetic and nonshared environmental influences on a trait. The trait terms are stable common factors, similar to the intercept term in a growth model, representing constant variation in a trait. The state term is represented by a simplex process, in which the value of the trait at time k is linearly related to the trait at time $k + 1$, with coefficient b generally less than 1.0. Over longer intervals spanning multiple simplex paths, therefore, the state-based stability declines exponentially as b_k.

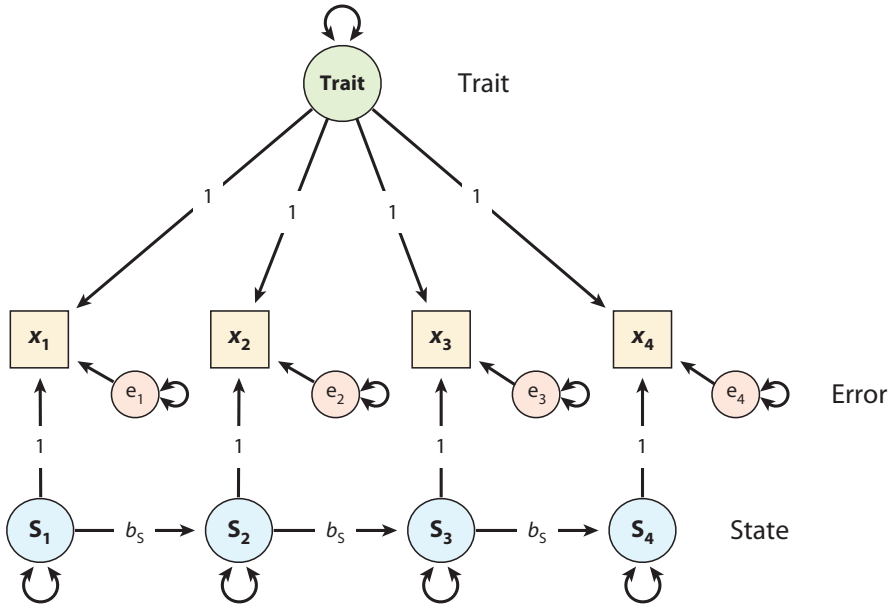

Figure 2

State-trait-error model representing stability in a phenotype as a sum of two components: a stable trait and a simplex state, which decays with time.

The stability of genetic and environmental influences on personality varies as a function of age and the interval between measurement occasions. Similar to most of the other effects reviewed in this article, they do not differ systematically in terms of the trait under study, at least within the domains of the major personality traits from the EPQ, FFM, and Minnesota-based systems. Because the most important determinant appears to be age, we organize our discussion into three sections: childhood and adolescence, early adulthood and mid-adulthood, and old age. We describe genetic and nonshared environmental correlations among measurement occasions for each study. Shared environment, as is typical, is mostly absent.

There are fewer genetically informed longitudinal studies of children than there are of adults. The youngest participants to be studied repeatedly are a sample of German twins described by Spengler et al. (2012), who describe self-report data from a sample of German children on a childhood version of the FFM at two measurement occasions at mean ages of 9 and 13 years. Genetic correlations between ages 9 and 13 were 0.72 for neuroticism, 0.32 for openness, and unity for the other FFM traits. Nonshared environmental correlations were 0.3 for neuroticism, 0.18 for extraversion, 0.28 for agreeableness, 0.31 for conscientiousness, and zero for openness. De Fruyt and colleagues (2006) examined Neuroticism and Extraversion scores in a sample of 548 Belgian children at a three-year interval starting at age 9. A correlations were close to unity, whereas E correlations ranged from 0.57 to 0.67. Gillespie et al. (2004) fit simplex models to EPQ scores obtained from a sample of Australian adolescent twins at ages 12, 14, and 16 years. Approximations of the longitudinal A and E correlations can be recovered from their simplex results. Genetic correlations ranged from 0.7 to unity, and E correlations ranged from 0.2 to 0.5. Bratko & Butkovic (2007) reported on a sample of Croatian twins tested with the EPQ at 17 and 21 years of age. The A correlations between the two ages were 0.87 for extraversion and 0.83 for neuroticism; the E correlations were 0.36 and 0.38.

Twin study: a comparison of the similarity of identical and fraternal twins, used to estimate genetic and environmental variance

Proceeding to early adulthood, the first extensive longitudinal twin study to be reported was that by McGue et al. (1993), which involved a small sample of 127 twin pairs from the Minnesota Twin Family Study who were tested with the MPQ twice at an average 10-year interval, at an

average age of 19.8 years at time 1 and 29.6 years at time 2. Results showed considerable phenotypic stability over time, which was largely attributable to genetic effects. For the three higher-order traits of PE, NE, and constraint, the genetic correlations between ages 20 and 30 were 0.81, 0.72, and 0.80, respectively. The corresponding E correlations were 0.30, 0.47, and 0.32. Viken et al. (1994) described a very large sample of Finnish twins who were measured twice for extraversion and neuroticism at an interval of 6 years. The twins varied in age at first testing from 18 to 48. Genetic correlations between measurements for both traits were at least 0.80 for the youngest (age 18) cohort and quickly rose to 1.0 thereafter. Nonshared environmental longitudinal correlations were approximately 0.3 for the 18-year-olds and slowly increased to approximately 0.5 for oldest cohort.

Blonigen et al. (2008) reported the MPQ scores of Minnesota twins tested at a 7-year interval, at ages 17 and 24. A third measurement occasion, at age 29, was added by Hopwood et al. 2011; we give only the results from the latter report. For the major MPQ dimensions of agentic and communal PE, NE, and constraint, A correlations increased from around 0.7 between ages 17 and 24 to 0.95 and greater between ages 24 to 29, but did not show a strong relation to length of interval. E correlations increased from approximately 0.35 between ages 17 and 24 to approximately 0.6 between ages 24 and 29.

Bleidorn et al. (2009) examined three measurement occasions of FFM data for the Bielefeld Longitudinal Study of Adult Twins (BiLSAT) sample in Germany, at mean ages 31, 36, and 40, although there was also considerable age variability within the cohort. Results were reported in the form of latent growth models. There was significant A and E variation in both the level (mean) and slope (also a source of stability) parameters for all of the FFM model traits and most of the facets; unfortunately, the residual terms uncorrelated with level were not ACE parameterized, making it impossible to compute the stability and instability of the A and E terms. Fortunately, Kandler et al. (2010) reported results from an expanded version of the BiLSAT sample. There were three measurement occasions in two cohorts, one at ages 23, 29, and 35 and the other at ages 39, 47, and 54. Genetic stability was nearly perfect throughout. Nonshared environmental stability was 0.37 at the earliest retest, between ages 23 and 29, and increased to 0.94 at the latest, between ages 47 and 54.

The literature also includes two interesting longitudinal studies conducted beginning in middle age. Johnson et al. (2005) described results for twins in late adulthood at two occasions, at average ages of 59 and 64. For the three broad domains of PE, NE, and constraint, the genetic correlations were 0.97, 1.0, and 0.93, respectively; the corresponding E correlations were 0.73, 0.71, and 0.64. Read et al. (2006) measured a sample of elderly twin pairs at three occasions with the EPQ, at ages 82, 84, and 86. Controls for mortality were included. All the genetic correlations were unity for both extraversion and neuroticism. E correlations ranged from 0.5 to 0.6 for extraversion and from 0.4 to 0.5 for neuroticism.

The results of studies of the stability of genetic and environmental variance in personality variables are difficult to interpret individually because of their joint dependence on the age at which the data were collected and the interval between the measurements. To examine the results in more depth, we recorded the A and E correlations between measurement occasions for all studies reporting more than one occasion of measurement for either neuroticism or extraversion. **Table 1** provides the recorded data. Notably omitted from the table is the study by Wray et al. (2007), who reported sophisticated analyses of genetic and environmental stability but used a design that included parents and children as well as twin pairs, making their sample impossible to characterize in terms of age. (All the genetic correlations were above 0.8, even at the 22-year interval, and did not show any systematic relationship to the length of the interval; E correlations ranged from 0.24 to 0.53, varying inversely with length of interval.) For studies reporting results separately for males and females, we computed the mean correlation. We included results for

Table 1 Genetic and unique environment correlations across time

Study	Trait	Age1	Interval	rA	rE
Bratko & Butkevic (2007)	E	17	4	0.87	0.36
	N	17	4	0.83	0.38
De Fruyt et al. (2006)	E	9	3	0.94	0.57
	N	9	3	1.00	0.67
Gillespie et al. (2004)	N	12	2	0.81	0.32
		12	4	0.74	0.24
		14	2	0.84	0.27
	E	12	2	0.88	0.32
		12	4	0.88	0.18
		14	2	0.96	0.39
Hopwood et al. (2011)	N	17	7	0.75	0.36
		17	12	0.86	0.32
		24	5	0.99	0.62
	E	17	7	0.73	0.38
		17	12	0.71	0.38
		24	5	0.96	0.57
Johnson et al. (2005)	NE	59	5	1.00	0.71
	PE	59	5	0.97	0.73
Kandler et al. (2010)	N	23	6	1.00	0.37
		23	12	1.00	0.25
		29	6	1.00	0.73
		41	7	1.00	0.58
		41	14	1.00	0.47
		48	7	1.00	0.94
	E	23	6	1.00	0.50
		23	12	1.00	0.28
		29	6	1.00	0.80
		41	7	1.00	0.82
		41	14	1.00	0.67
		48	7	1.00	0.89
Read et al. (2006)	N	82	2	1.00	0.48
		82	2	1.00	0.44
		84	2	1.00	0.40
	E	82	2	1.00	0.51
		82	2	1.00	0.57
		84	2	1.00	0.54
Spengler et al. (2012)	N	9	3	0.72	0.30
		9	3	1.00	0.18
Viken et al. (1994)	N	21	6	0.83	0.25
		27	6	1.00	0.35
		33	6	0.84	0.47
		39	6	1.00	0.35

(*Continued*)

Table 1 (*Continued*)

Study	Trait	Age1	Interval	rA	rE
	E	45	6	1.00	0.48
		51	6	1.00	0.47
		21	6	0.87	0.35
		27	6	1.00	0.44
		33	6	1.00	0.51
		39	6	1.00	0.50
		45	6	1.00	0.52
		51	6	1.00	0.48
McGue et al. (1993)	NE	20	10	0.72	0.47
	PE	20	10	0.81	0.30
Wray et al. (2007)	N	—	9	0.91	0.53
			19	0.93	0.38
			22	0.95	0.24
			10	0.95	0.44
			13	0.88	0.42
			3	0.82	0.48

Abbreviations: Age1, age at assessment occasion one; N, neuroticism; E, extraversion; NE, negative emotionality; PE, positive emotionality; rA, genetic correlation; rE, nonshared environment correlation.

PE with extraversion and NE with neuroticism. Kandler (2012) conducted a similar analysis of personality stability as a function of age but not interval between measurements. Our results suggest that it is difficult to understand the results of one without including the other.

Figure 3 illustrates the univariate relationships of age and interval as predictors of stability for A and C. For A (**Figure 3*a***), stability rises quickly through early adulthood, reaching 1.0 by the mid-twenties. The stability of E is lower in general, also rises through middle age, and appears to drop off in old age. Genetic stability does not show a clear relationship with the interval

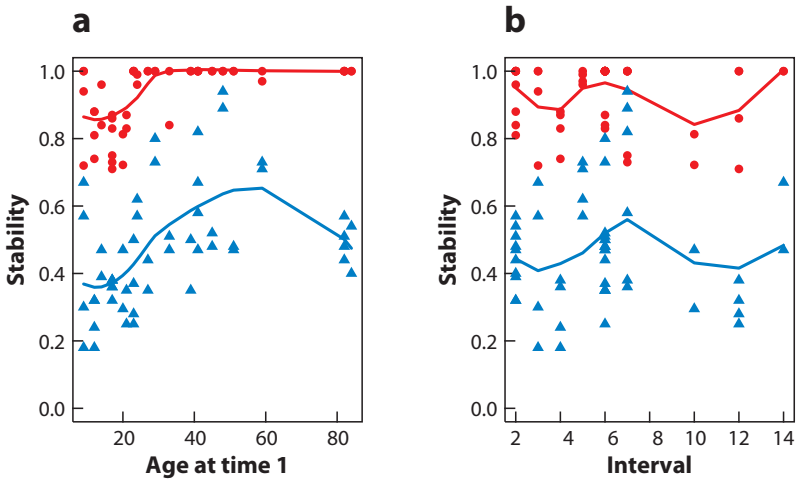

Figure 3

Univariate relation between genetic (*red circles*) and shared environmental (*blue triangles*) test-retest correlations and (*a*) age at first measurement and (*b*) interval between testings.

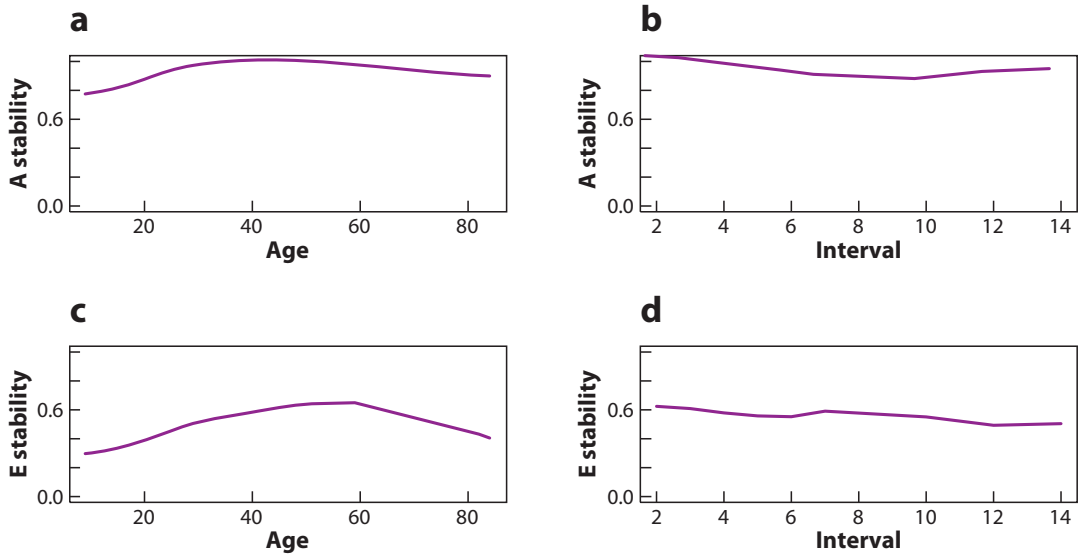

Figure 4

(*a*,*b*) Results of generalized additive model of genetic and (*c*,*d*) nonshared environmental stabilities as a function of age (*a*,*c*) and (*b*,*d*) interval between measurements. Abbreviations: A, genetic effects; E, nonshared environmental effects.

between measurements, whereas E stability appears to diminish at longer intervals. The univariate relationships between age and interval on the one hand and stability on the other are difficult to discern in the univariate plots, however, because age and interval are themselves correlated. Studies starting with younger participants have longer intervals of measurement. We therefore analyzed the joint effect of age and interval.

For several reasons, we analyzed these results as nonparametrically as possible. First, the observations are not independent, with multiple coefficients reported from individual studies. Second, correlation coefficients as an outcome variable are not normally distributed, and many of these are closely bounded by 1.0. Third, there is good reason to expect the relationship between stability and both age and interval to be nonlinear. For all of these reasons, we decided to fit a generalized additive model (GAM) (Hastie & Tibshirani 1990) to the genetic and nonshared environmental stabilities. A GAM model fits nonlinear smooth functions to the relations between multiple predictors and the outcome; the effect of each predictor is conditional on the other predictors. These smooth functions can then be added together, in a nonlinear smoothing-based analog of multiple regressions, to produce a nonlinear multivariate predictor of the outcome.

Figure 4 shows the results. There were no differences in stability between neuroticism and extraversion, so the results are reported together. Genetic stability increases rapidly through the mid-twenties and remains close to unity throughout the life span, decreasing only slightly in old age; no relationship between genetic stability and the interval between the measurements is apparent. In contrast, nonshared environmental stability increases gradually through midlife and then drops off substantially in old age. Nonshared environmental stability also shows a strong relationship with the interval between the measurements. After early adulthood, genetic influences on personality are nearly perfectly stable and time independent; nonshared environmental influences decrease slowly over time with exponentially decreasing effects and become unstable in old age.

Consider a hypothetical pair of identical twins whose personalities develop throughout the life span. In the usual absence of shared environmental effects, their pair-average score on personality traits is highly, even perfectly, stable relative to the average of other pairs, largely as an expression

of their genetic endowment. At the same time, within-pair differences between twins show environmental influences that come and go, systematically but temporarily, from childhood through late middle age. At any given point in time, something might happen to make one member of the pair more extraverted or more neurotic; then, as time goes by, the within-pair difference decays, and the twins return to their genetically influenced mean. Finally, in old age, new differential processes appear to be established that make within-pair differences much less stable over time.

A PHENOTYPIC NULL HYPOTHESIS

Loehlin & Martin (2013), at the end of their paper that marks the conclusion of Loehlin's long investigation of personality structure in genetic and environmental domains, remark that personality appears to be a "phenotypic process." What do they mean by that? In some sense, of course, everything psychological is a phenotypic process, in that we observe it at the level of the phenotype. But not everything is phenotypic in the same way. Suppose we observe a patient with Huntington's disease, exhibiting the writhing choreiform movements and cognitive disabilities that are characteristic of the disorder. Those symptoms are phenotypic, but we would not say that Huntington's disease itself is a phenotypic process, and we would not seek psychological or social explanations of the co-occurrence of choreiform movement and dementia. Why not? Because we know that a circumscribed, theoretically coherent explanation of the Huntington's phenotype exists at the genetic level. The Huntington's genotype explains the Huntington's phenotype. We could say, following Meehl (1977), that Huntington's disease has a specific genetic etiology.

By way of contrast, consider marital status. Divorce is heritable (Johnson et al. 2004), but do we really expect that intensive twin studies of marital processes will lead us to a genetic explanation of divorce? Presumably not, although twin studies can be informative about marriage in other important ways, especially as quasi-experimental tests of causal hypotheses, in the manner of the validity studies reviewed above. For marriage and divorce, we recognize that the observed phenotypic structure as it exists in our culture is the result of psychological- and social-level (that is, phenotypic) forces. The point is not that they are environmental as opposed to genetic; indeed, as we cannot emphasize enough, marriage, divorce, and whatever may cause them are just as heritable as anything else. The heritability of marriage is a by-product of the universal, nonspecific, genetic pull on everything, not an indication that divorce is a biological process awaiting genetic analysis. Marriage and divorce are heritable, but they do not have a specific genetic etiology. For marriage and divorce, genotypic variation is phenotypic variation observed at a different level of analysis.

The laws of behavior genetics are not actual laws, and calling them that may have led to some misunderstandings of what was originally intended. In particular, although the First Law of Behavior Genetics has sometimes been considered an endorsement of a hereditarian view of behavior (e.g., Pinker 2003), the universality of heritability is best interpreted as a *reductio ad absurdum* of these very distinctions—as a way of observing that the endeavor of figuring out how genetic or environmental a trait is, let alone declaring that it is exclusively one or the other, is pointless. The laws of behavior genetics are collectively a null hypothesis and were described as such the first time they were mentioned (Turkheimer & Gottesman 1991). An updated version of the laws taking these multivariate findings into account, which we propose to call the phenotypic null hypothesis of behavior genetics, can be stated simply as follows: All traits are heritable, and the multivariate structure of the biometric components of behavior does not differ from the phenotypic structure.

The phenotypic null hypothesis is a statement of the simplest and most general reason that a phenotype would be heritable, the simplest version of what we already know, a standard against which the novelty of new hypotheses can be judged. In its original form, the First Law was an acknowledgment that we already knew on the basis of univariate behavior genetics: Behavior is

heritable; $h^2 > 0$; and as such there is little point in either demonstrating heritability over and over or trying to argue it away. The phenotypic null hypothesis integrates what we have learned from multivariate behavior genetics: Until demonstrated otherwise, complex heritable behavioral traits should be the result of psychological processes defined at a high level of analysis, rather than at the level of genes or neurons. Although such complex traits are never independent of genetic variation, they cannot be defined by genetic processes. Genes and behavior are a single entity, a single organism observed at two levels of analysis. Some traits are better understood using low-level concepts (genes, neurons, structures), whereas others require high-level constructs (organizations, algorithms, beliefs). Personality, for the first time in the history of genetically informed scientific psychology, has turned out to be a clear instance of the latter case.

Linkage analysis: association between DNA and a phenotype as they are transmitted within a family

The remarkable finding that the multivariate structure of personality is not detectably different across biometric domains is actually a direct prediction from the phenotypic null hypothesis. "The puzzle of parallel structures" (McCrae et al. 2001, p. 515) is a puzzle only because of our ingrained expectation that the heritability of the traits of the FFM implies the existence of genetic mechanisms, which somehow are more biological than the psychological and phenotypic patterning of behavior that we actually observe. The phenotypic null hypothesis can be considered in another way in the analysis of genetic and environmental stability in personality. Nonshared environmental effects on personality are modest but detectable, and they exist in time: Something happens to make one twin more extraverted than the other at a measurement occasion, and at some later point in time the same twin will still be more extraverted but less so, with the difference continuing to decay as more time passes. Genetic effects do not operate in this way. If a pair of identical twins is, on average, more extraverted than another pair, they will tend to remain so regardless of the length of time that passes before they are assessed again. Genetic differences in adults are almost perfectly stable, as one would expect if genes were exerting a steady but nonspecific pull on the phenotype.

The phenotypic null hypothesis plays out somewhat differently in validity studies. We suggest that, when studying relations between personality variables and differences in other domains of behavior, our scientific hypotheses are usually phenotypic in nature. When we hypothesize that an extraverted personality predisposes individuals to externalizing problems, for example, the causal hypothesis itself is from extraverted phenotype to externalizing phenotype, and the possibility that the phenotypic association between the two is the result of a shared genetic background is an alternative to the causal hypothesis. It is not surprising, then, that this is the area in which the phenotypic null hypothesis is most frequently rejected, because genuine causal relations between one domain and another are very hard to find. Most of the time, the reason differences in behavior are correlated across domains is that they share a general unsystematic genetic background. That is not the same as saying that there is a specific genetic mechanism that links them; the most we can say is that their nonspecific genetic backgrounds overlap.

GENOMICS OF PERSONALITY

As was the case for both the early experimental research on temperament in dogs and the large-scale twin studies of the past century, personality has played a central role in our understanding of the genetics of behavior as the sequencing of the human genome has been completed and it has become possible to assess genomic variation directly in the DNA. The still-brief history of molecular genetic studies of personality comes in three distinct phases: (*a*) an anticipatory phase that looked to estimates of heritability from the quantitative genetics of personality as a guide to expectations for the success of molecular genetics (Martin et al. 1997), (*b*) a period of considerable empirical optimism as early linkage and association studies produced multiple significant effects

for various personality traits, and (*c*) an extended period of frustration as most of the reported discoveries either turned out to be very small or failed to replicate at all.

Candidate gene association: a statistical association between a gene and a phenotype, usually across families

Genomewide association study (GWAS): a relatively inexpensive method to assess associations with as many as one million SNPs and a phenotype in large samples

Single-nucleotide polymorphism (SNP): a single unit of DNA that takes only two values across people

Two relatively simple methods prevailed in the early days of the molecular genetics of personality. Linkage analysis is a technique that looks for within-family cosegregation between a trait of interest and a genetic marker. Candidate gene association studies are even simpler: They are no more than correlational analyses of relations between identified genes and either categorical or continuous outcomes. Linkage and association studies have complementary strengths and weaknesses. Association studies are vulnerable to population stratification, meaning that there are many possible causal pathways that could explain a correlation between a gene and a personality trait, many of which are obviously spurious (Hamer & Sirota 2000). If members of one ethnic group are more extraverted than members of another, then in an ethnically mixed sample any gene that differs between the groups for any reason will be correlated with extraversion (Turkheimer 2012). Linkage studies similarly control for population stratification by focusing on within-family relationships that cannot be accounted for by environmental or cultural confounds varying at the level of family or culture. The inferential benefit comes with a price in statistical power because linkage studies require not only multiple family members but also family members who differ on the outcome of interest.

Despite these difficulties, early reports of linkage and association suggested that the optimism about the potential for a molecular genetics of personality would be borne out. Several reports of associations between the D4 receptor gene (*DRD4*) and sensation seeking (Benjamin et al. 1996), and between the serotonin transporter gene (*SLC6A4*) and anxiety-related traits (Lesch et al. 1996), were met with considerable enthusiasm and cited widely. As the literature evolved, however, it became more and more difficult to establish a clear replication of the early findings, at least at anything close to the large effect sizes that had been reported in the early papers. A series of meta-analyses conducted by Munafò et al. (2003, 2008, 2009) established quite clearly that the associations, if they differed from zero at all, were vanishingly small. The most recent review (Munafò & Flint 2011, p. 395) is very gloomy in outlook: "The first candidate gene studies of human personality promised much but, in the fifteen years since their publication, have delivered little in the way of clear evidence for the contribution of specific genetic variants to observed variation in personality traits."

Ironically, the most recent developments in genomic technology have worked against confidence in the magnitude of the findings from linkage and candidate gene association studies. The rapid development of genomewide association studies (GWASs) has enabled the inexpensive assessment of variation of initially hundreds of thousands and, later, millions of single-nucleotide polymorphisms (SNPs). GWASs enabled the testing of associations with hundreds of thousands of loci at low cost, but they introduced obvious problems of multiple hypothesis testing. The broader genomics community responded by abandoning theoretically driven selection of candidate gene hypotheses in favor of extremely stringent ($p < 10^{-8}$) hurdles for genomewide significance, which, in the face of ever-smaller effect sizes, demanded larger and larger samples (Lander & Schork 1994).

Although GWASs have produced some notable successes in medicine [and it is fair to say that the jury is still out on neuropsychiatry (Visscher & Montgomery 2009)], in personality it is difficult to point to any successes at all from GWASs. It is not that statistical significance, even genomewide statistical significance, has never been achieved. Indeed, most GWASs report one or two associations at or close to genomewide significance, which then are not replicated in the next study. No GWAS of personality variables has ever reported an association accounting for as much as 0.05% of the variance. None of the classic loci from the early era of candidate gene association studies have ever surpassed or even approached genomewide significance.

In the most comprehensive GWAS of personality conducted to date, de Moor et al. (2010) combined results from 10 independent samples constituting a total of 17,375 adults, and withheld five additional samples, with 3,294 adults, for replication. All participants were of European ancestry and had been administered the NEO-PI, and information on 2.4 million SNPs was available. These authors calculated results in the individual discovery samples, combined using meta-analytic procedures and replicated in the withheld samples. Two SNPs showed genomewide association with openness and one with conscientiousness. Each of the three SNPs accounted for a little more than 0.2% of the variation in the corresponding personality trait. The effects did not replicate in the withheld samples; no external replication attempts have yet been reported.

The difficulties encountered in the molecular genetics of personality are a reflection of the phenotypic null hypothesis operating at a genomic level of organization. The question of whether there were associations to be found between individual genes or SNPs and variation in personality was settled on the day it was agreed that identical twins were more correlated for neuroticism than were fraternal twins. If one accepts that neuroticism is heritable, what mechanisms are available other than the cumulative effects of genes at multiple loci? However, the causal structure, as opposed to the mere existence, of molecular genetic associations with personality is exactly as would be predicted by the phenotypic null hypothesis. The more similar people are in genotype, the more similar they are in personality, but genotypic similarity appears to be carried across many—by current indications, uncountable—genes with effects that are both tiny and unsystematic, beyond their cumulative effect of making people who share them similar in general.

CONCLUSION

Null hypotheses cannot be confirmed, but the conclusion of this review is that in the genetics of personality, a paradoxical outcome that has been looming for a long time has finally come to pass: Personality is heritable, but it has no genetic mechanism. The prospect of this outcome has haunted the nature-nurture debate from its inception, as both sides of the old debate were led to a dead end of thinking that the point of the debate was to evaluate the separate effects of genes and environment. It became clear long ago that neither genes nor environment could be discounted for anything important, a conclusion that stalled the discussion either in intransigent hereditarian and environmentalist positions or in an unsatisfying interactionist middle ground.

Although the search for genetic mechanisms of human personality, in our view, will never bear fruit, it is nevertheless possible to construct a genetically informed phenotypic science of behavior. Behavior genetic methods will not provide a mechanism in such a science; instead, they will provide a means of establishing quasi-experimental control over familial associations that otherwise confound associations among human variables in nonexperimental settings. The heritability of personality has one important consequence that cannot be restated often enough: Uncontrolled correlations between the behaviors of genetically related individuals are not necessarily causal, let alone environmental. If extraverted mothers have extraverted children, it is not necessarily the case that the children are learning to be extraverted by modeling their parents' behavior. This caveat remains in effect no matter how the genetics of extraversion actually works; it remains in effect if there is no more of a genetic mechanism for extraversion than there is for divorce. Genetically informed research designs can partially, imperfectly, control for the genetic and shared environmental confounds that otherwise cloud causal interpretation of associations like these, and they have been extraordinarily successful at doing so. The quantification of heritability itself is unimportant in such analyses, except as a node in statistical models that control for genetic pathways in nonexperimental studies.

Perhaps because the results of GWASs of personality appear so bleak, the personality field has largely avoided the most common conclusion reached on the basis of the almost-as-discouraging results that have emerged from the molecular genetics of other behavioral traits like intelligence or psychopathology. GWASs, it is said, have demonstrated that the effects of individual genes are universally small; even the largest accounts for less than 1% of the variance. Therefore, we will need ever-larger studies, consortia of studies, and meta-analyses of consortia to detect the vanishingly small effects of individual genes. We are skeptical that this strategy will be successful for personality in the long run. Can one point to a field of science that has been successful by stringing together the multiple effects of such tiny associations? The phenotypic null hypothesis suggests that the foundational idea that there are individual causal genetic variants for personality, however small, is itself flawed. Except in the weakest statistical sense, there actually is not a large set of neuroticism genes, each with small effect; there is merely a nonspecific genetic background to phenotypic neuroticism, and to its phenotypic causes and effects.

When Galton first formulated the nature-nurture debate in the nineteenth century, the alternative to "genetic" was supposed to be "environmental." That classical version of the behavior genetic analysis of personality has finally reached a clear conclusion. Both genes and environments matter, but neither genetic nor environmental effects can be broken down into discrete and specifiable mechanisms at a lower level of analysis. The establishment of genetic and environmental variance in personality has answered important questions, but as the genes-versus-environment version of the debate has reached its end, it has turned out that another question, about the existence of lower-level mechanisms for observed phenotypic behavior, constituted a large part of what we wanted to know all along. In observing, again and again, the heritability and environmentality of behavior in general and personality in particular, we have assumed that the causal (or at least the explanatory) arrows must be directed from the bottom up. The phenotypic null hypothesis suggests that the explanatory direction is exactly the reverse: Phenotypic variation explains the genetic structure of behavior. If the failure to reject the phenotypic null hypothesis for the genetics of personality represents a victory for any particular mode of explanation, the winner is not naïve environmentalism but rather biologically informed psychological explanation; the loser is not genetics but rather poorly informed and superficial biologism.

SUMMARY POINTS

1. All personality traits are heritable, and equally so. To the limited extent it is possible to specify numerical values of heritability at all, all personality traits are heritable at about $h^2 \approx 0.4$. Narrow traits in sufficiently large samples sometimes show significant differences, but these do not replicate from one study to another.
2. The heritability of personality exists at all levels of its hierarchical structure. Personality items are heritable, narrow facets are heritable, the traits of the FFM of personality and other systems are heritable, and high-order traits are heritable. The only systematic differences among the levels involve the progressive elimination of measurement error.
3. The multivariate structures of the three genetic and environmental biometric components of personality do not differ from each other and, therefore, do not differ from the phenotypic structure of personality that they jointly compose.
4. Most observed associations among personality differences and other variables are a combination of a noncausal shared genetic background and a smaller, plausibly causal phenotypic remainder that operate within pairs of identical twins raised together.

5. The developmental structure of phenotypic personality stability as a function of age is a combination of (*a*) genetic differences that become nearly perfectly stable in early adulthood and do not decay over time and (*b*) environmental differences that also become more stable, but are generally less so, become more unstable in late life and decay slowly over time.

6. DNA-based studies have shown that the heritability of human personality is based on the accumulated action of a very large number of genes. Attempts to specify individual genes causing differences in personality traits have not been successful.

FUTURE ISSUES

1. Larger and larger GWASs are being conducted, allowing researchers to detect smaller and smaller associations between SNPs and personality traits with genomewide significance. Whether such associations, which will almost certainly be smaller than $r = 0.02$, will have meaningful psychological or biological content remains to be determined.
2. Genomic technology is proceeding rapidly. In particular, it will soon be possible to obtain the full genetic sequence on large numbers of people, which will provide more information than can be obtained from SNPs and GWASs. Whether full-genome sequencing will provide a more detailed account of genetic mechanisms underlying human personality remains to be seen.
3. Some new technologies, such as genomic complex-trait analysis, focus more on predicting personality from the full genome rather than on finding individual genes that are associated with specific traits. Currently, the ability to predict personality from genomic data is quite low. It is not known how much higher it can get.
4. If it ever became possible to predict personality from genomic data alone, there would be profound ethical issues involved in the use of the data for reproductive decision making or scientific purposes.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Benjamin J, Li L, Patterson C, Greenberg BD, Murphy DL, Hamer DH. 1996. Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nat. Genet.* 12:81–84

Bleidorn W, Kandler C, Riemann R, Angleitner A, Spinath FM. 2009. Patterns and sources of adult personality development: growth curve analyses of the NEO PI-R scales in a longitudinal twin study. *J. Personal. Soc. Psychol.* 97:142–55

Blonigen DM, Carlson MD, Hicks BM, Krueger RF, Iacono WG. 2008. Stability and change in personality traits from late adolescence to early adulthood: a longitudinal twin study. *J. Personal.* 76:229–266

Bouchard TJ Jr, Loehlin JC. 2001. Genes, evolution, and personality. *Behav. Genet.* 31:243–73

Bratko D, Butkovic A. 2007. Stability of genetic and environmental effects from adolescence to young adulthood: results of Croatian longitudinal twin study of personality. *Twin Res. Hum. Genet.* 10:151–57
Carey G, DiLalla DL. 1994. Personality and psychopathology: genetic perspectives. *J. Abnorm. Psychol.* 103:32–43
Carey G, Goldsmith HH, Tellegen A, Gottesman II. 1978. Genetics and personality inventories: the limits of replication with twin data. *Behav. Genet.* 8:299–313
Caspi A, Roberts BW, Shiner RL. 2005. Personality development: stability and change. *Annu. Rev. Psychol.* 56:453–84
Cattell RB, Rao DC, Schuerger JM, Vaughan DS. 1981. Unitary personality source traits analyzed for heritability. *Hum. Hered.* 31(5):261–75
Charney E. 2012. Behavior genetics and postgenomics. *Behav. Brain Sci.* 35:331–58
De Fruyt F, Bartels M, Van Leeuwen KG, De Clercq B, Decuyper M, Mervielde I. 2006. Five types of personality continuity in childhood and adolescence. *J. Personal. Soc. Psychol.* 91(3):538–52
de Moor MH, Costa PT, Terracciano A, Krueger RF, De Geus EJC, et al. 2010. Meta-analysis of genome-wide association studies for personality. *Mol. Psychiatry* 17:337–49
DeYoung CG. 2006. Higher-order factors of the Big Five in a multi-informant sample. *J. Personal. Soc. Psychol.* 91:1138–51
Digman JM. 1997. Higher-order factors of the Big Five. *J. Personal. Soc. Psychol.* 73:1246–56
Eysenck HJ. 1990. Genetic and environmental contributions to individual differences: the three major dimensions of personality. *J. Personal.* 58(1):245–61
Fanous A, Gardner CO, Prescott CA, Cancro R, Kendler KS. 2002. Neuroticism, major depression and gender: a population-based twin study. *Psychol. Med.* 32:719–28
Fuller JL. 1960. Behavior genetics. *Annu. Rev. Psychol.* 11:41–70
Gillespie NA, Evans DE, Wright MM, Martin NG. 2004. Genetic simplex modeling of Eysenck's dimensions of personality in a sample of young Australian twins. *Twin Res.* 7:637–48
Hamer D, Sirota L. 2000. Beware the chopsticks gene. *Mol. Psychiatry* 5:11–13
Hastie T, Tibshirani R. 1990. *Monographs on Statistics and Applied Probability*, vol. 43: *Generalized Additive Models*. New York: Chapman & Hall
Heath AC, Cloninger CR, Martin NG. 1994. Testing a model for the genetic structure of personality: a comparison of the personality systems of Cloninger and Eysenck. *J. Personal. Soc. Psychol.* 66:762–75
Heath AC, Kessler RC, Neale MC, Hewitt JK, Eaves LJ, Kendler KS. 1993. Testing hypotheses about direction of causation using cross-sectional family data. *Behav. Genet.* 23:29–50
Hopwood CJ, Donnellan MB, Blonigen DM, Krueger RF, McGue M, et al. 2011. Genetic and environmental influences on personality trait stability and growth during the transition to adulthood: a three-wave longitudinal study. *J. Personal. Soc. Psychol.* 100:545–56
Horn JM, Plomin R, Rosenman R. 1976. Heritability of personality traits in adult male twins. *Behav. Genet.* 6:17–30
Jang KL, Livesley WJ, Ando J, Yamagata S, Suzuki A, et al. 2006. Behavioral genetics of the higher-order factors of the Big Five. *Personal. Individ. Differ.* 41:261–72
Jang KL, Livesley WJ, Angleitner A, Riemann R, Vernon PA. 2002. Genetic and environmental influences on the covariance of facets defining the domains of the five-factor model of personality. *Personal. Individ. Differ.* 33:83–101
Jang KL, McCrae RR, Angleitner A, Riemann R, Livesley WJ. 1998. Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. *J. Personal. Soc. Psychol.* 74:1556–65
Johnson W, McGue M, Krueger RF. 2005. Personality stability in late adulthood: a behavioral genetic analysis. *J. Personal.* 73:523–52
Johnson W, McGue M, Krueger RF, Bouchard TJ Jr. 2004. Marriage and personality: a genetic analysis. *J. Personal. Soc. Psychol.* 86:285–94
Joseph J. 2004. *The Gene Illusion: Genetic Research in Psychiatry and Psychology Under the Microscope*. New York: Algora
Kandler C. 2012. Nature and nurture in personality development. The case of neuroticism and extraversion. *Curr. Dir. Psychol. Sci.* 21(5):290–96

Kandler C, Bleidorn W, Riemann R, Spinath FM, Thiel W, Angleitner A. 2010. Sources of cumulative continuity in personality: a longitudinal multiple-rater twin study. *J. Personal. Soc. Psychol.* 98:995–1008
Kandler C, Riemann R, Spinath FM, Angleitner A. 2010. Sources of variance in personality facets: a multiple-rater twin study of self-peer, peer-peer, and self-self (dis)agreement. *J. Personal.* 78:1565–94
Kendler KS, Gatz M, Gardner CO, Pedersen NL. 2006. Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch. Gen. Psychiatry* 63:1113–20
Kenny DA, Zautra A. 1995. The trait-state-error model for multiwave data. *J. Consult. Clin. Psychol.* 63:52–59
Klump KL, McGue M, Iacono WG. 2002. Genetic relationships between personality and eating attitudes and behaviors. *J. Abnorm. Psychol.* 111:380–89
Krueger RF. 2000. Phenotypic, genetic, and nonshared environmental parallels in the structure of personality: a view from the Multidimensional Personality Questionnaire. *J. Personal. Soc. Psychol.* 79:1057–67
Krueger RF, Johnson W. 2008. Behavioral genetics and personality: a new look at the integration of nature and nurture. *Handb. Personal. Theory Res.* 3:287–310
Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. *Science* 264:2037–48
Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, et al. 1996. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274:1527–31
Livesley WJ, Jang KL, Vernon PA. 1998. Phenotypic and genetic structure of traits delineating personality disorder. *Arch. Gen. Psychiatry* 55:941–48
Loehlin JC. 1982. Are personality traits differentially heritable? *Behav. Genet.* 12:417–28
Loehlin JC. 1978. Are CPI scales differently heritable? How good is the evidence? *Behav. Genet.* 8:381–82
Loehlin JC. 1985. Fitting heredity-environment models jointly to twin and adoption data from the California Psychological Inventory. *Behav. Genet.* 15:199–221
Loehlin JC. 1987. Heredity, environment, and the structure of the California Psychological Inventory. *Multivar. Behav. Res.* 22:137–48
Loehlin JC. 2011. Genetic and environmental structures of personality: a cluster-analysis approach. *Personal. Individ. Differ.* 51:662–66
Loehlin JC. 2012. The differential heritability of personality item clusters. *Behav. Genet.* 42:500–7
Loehlin JC, Martin NG. 2013. General and supplementary factors of personality in genetic and environmental correlation matrices. *Personal. Individ. Differ.* 54:761–66
Loehlin JC, McCrae RR, Costa PT, John OP. 1998. Heritabilities of common and measure-specific components of the Big Five personality factors. *J. Res. Personal.* 32:431–53
Loehlin JC, Neiderhiser JM, Reiss D. 2003. The behavior genetics of personality and the NEAD study. *J. Res. Personal.* 37(5):373–87
Loehlin JC, Nichols RC. 1976. *Heredity, Environment, and Personality: A Study of 850 Sets of Twins*. Austin: Univ. Tex. Press
Lykken DT. 1982. Research with twins: the concept of emergenesis. *Psychophysiology* 19:361–72
Martinl N, Boomsma D, Machin G. 1997. A twin-pronged attack on complex traits. *Nat. Genet.* 17:387–92
McGue M, Bacon S, Lykken DT. 1993. Personality stability and change in early adulthood: a behavioral genetic analysis. *Dev. Psychol.* 29:96–109
McCrae RR, Jang KL, Livesley WJ, Riemann R, Angleitner A. 2001. Sources of structure: genetic, environmental, and artifactual influences on the covariation of personality traits. *J. Personal.* 69:511–35
Meehl PE. 1967. Theory-testing in psychology and physics: a methodological paradox. *Philos. Sci.* 34:103–15
Meehl PE. 1977. Specific etiology and other forms of strong influence: some quantitative meanings. *J. Med. Philos.* 2:33–53
Munafò MR, Clark TG, Moore LR, Payne E, Walton R, Flint J. 2003. Genetic polymorphisms and personality in healthy adults: a systematic review and meta-analysis. *Mol. Psychiatry* 8:471–84
Munafò MR, Clark T, Flint J. 2004. Does measurement instrument moderate the association between the serotonin transporter gene and anxiety-related personality traits? A meta-analysis. *Mol. Psychiatry* 10:415–19
Munafò MR, Brown SM, Hariri AR. 2008. Serotonin transporter (5-HTTLPR) genotype and amygdala activation: a meta-analysis. *Biol. Psychiatry* 63:852–57
Munafò MR, Flint J. 2011. Dissecting the genetic architecture of human personality. *Trends Cogn. Sci.* 15:395–400

Munafò MR, Freimer NB, Ng W, Ophoff R, Veijola J, et al. 2009. 5-HTTLPR genotype and anxiety-related personality traits: a meta-analysis and new data. *Am. J. Med. Genet. B* 150:271–81

Pinker S. 2003. *The Blank Slate: The Modern Denial of Human Nature*. New York: Penguin

Pettersson E, Turkheimer E, Horn EE, Menatti AR. 2012. The general factor of personality and evaluation. *Eur. J. Personal.* 26:292–302

Plomin R, Caspi A. 1990. Behavioral genetics and personality. *Handb. Personal. Theory Res.* 2:251–76

Read S, Vogler GP, Pedersen NL, Johansson B. 2006. Stability and change in genetic and environmental components of personality in old age. *Personal. Individ. Differ.* 40:1637–47

Rushton JP, Bons TA, Hur YM. 2008. The genetics and evolution of the general factor of personality. *J. Res. Personal.* 42:1173–85

Spengler M, Gottschling J, Spinath FM. 2012. Personality in childhood—a longitudinal behavior genetic approach. *Personal. Individ. Differ.* 53:411–16

Tukey JW. 1954. Causation, regression, and path analysis. In *Statistics and Mathematics in Biology*, ed. O Kempthorne, TA Bancroft, JW Gowen, JL Lush, pp. 35–66. Ames: Iowa State Univ. Press

Thompson WR, Wilde GJS. 1973. Behavior genetics. In *Handbook of General Psychology*, ed. BB Wolman, pp. 206–29. Englewood Cliffs, NJ: Prentice-Hall

Turkheimer E. 1998. Heritability and biological explanation. *Psychol. Rev.* 105:782–91

Turkheimer E. 2000. Three laws of behavior genetics and what they mean. *Curr. Dir. Psychol. Sci.* 9:160–64

Turkheimer E. 2011. Still missing. *Res. Hum. Dev.* 8:227–41

Turkheimer E. 2012. Genome wide association studies of behavior are social science. In *Philosophy of Behavioral Biology*, ed. KS Plaisance, TAC Reydon, pp. 43–64. New York: Springer

Turkheimer E, Gottesman II. 1991. Is $H^2 = 0$ a null hypothesis anymore? *Behav. Brain Sci.* 14:410–11

Turkheimer E, Harden KP. 2013. Behavior genetic research methods: testing quasi-causal hypotheses using multivariate twin data. In *Handbook of Research Methods in Personality and Social Psychology*, ed. HT Reis, CM Judd. Cambridge Univ. Press. 2nd ed. In press

Turkheimer E, Waldron M. 2000. Nonshared environment: a theoretical, methodological, and quantitative review. *Psychol. Bull.* 126:78–108

Viken RJ, Rose RJ, Kaprio J, Koskenvuo M. 1994. A developmental genetic analysis of adult personality: extraversion and neuroticism from 18 to 59 years of age. *J. Personal. Soc. Psychol.* 66:722–22

Visscher PM, Montgomery GW. 2009. Genome-wide association studies and human disease. *JAMA* 302:2028–29

Wray NR, Birley AJ, Sullivan PF, Visscher PM, Martin NG. 2007. Genetic and phenotypic stability of measures of neuroticism over 22 years. *Twin Res. Hum. Genet.* 10:695–702

Yamagata S, Suzuki A, Ando J, Ono Y, Kijima N, et al. 2006. Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. *J. Personal. Soc. Psychol.* 90:987–98

Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, et al. 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* 43:519–25

Annual Review of Psychology

Volume 64, 2013

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at http://psych.AnnualReviews.org/errata.shtml