A More Accurate Estimate of Heritability

David T. Lykken†

Department of Psychology, University of Minnesota, Minneapolis, United States of America

All psychological and psychophysiological traits vary, often widely, about their stable set-point values, due to transitory environmental influences. Because it is this stable set-point that embodies the genetically determined component of the trait, twin and family data based on one-time trait measurements must underestimate true trait heritability. The means of multiple measurements, taken months or years apart, then correlated within pairs of monozygotic twins, would yield an accurate estimate of the broad heritability of the set-point value, but such data are rare and expensive. Given just two measurements sufficiently far apart, the cross-twin cross-time correlation (R_{CT}), divided by the retest or within-twin cross-time correlation (R_{WT}), provides a valid estimate of set-point heritability. This article examines data from young and middle-aged twins who were tested twice, 3 or more years apart, on heart rate and blood pressure, personality traits, self-rating items, occupational and recreational interests, as well as on Wechsler Intelligence Scales. In every case, the disattenuated R_{CT} revealed substantially higher heritability than indicated by correlations based on single measurements.

One of the attractions that psychological research still has for the young is that there remain so many easy and obvious improvements to be made in its theories and methods (e.g., Lykken, 1991). One example is the almost universal but erroneous assumption that a single measure of a trait constitutes the lasting sum or product of genetic and environmental influences and that, apart from an inevitable but (usually) small amount of measurement error, that single measure reflects the trait's stable value or intensity. For example, when McGue, Bacon, and Lykken (1992) administered the Multidimensional Personality Questionnaire (MPQ: Tellegen & Waller, 1994) twice, to twin pairs who were approximately 20 and 30 years old, on the two occasions we found mean retest correlations of only .53 for the 11 scales and .59 for the three factors. We concluded that there had been considerable personality change during that eventful 10-year period and, since the monozygotic (MZ) twins gave within-pair correlations on both occasions of about .51 for the scales and .55 for the factors, we also concluded that the evolving trait values had remained about 50% heritable.

But it is obvious that most traits of interest — for example, blood pressure (BP), EEG alpha activity, aggressiveness, subjective wellbeing, impulsiveness, anxiety, recreational interests, sexual interests, even IQ and mental vigor — vary in amplitude from time to time about their stable or *set-point* value (Carey, 2003; Lykken & Tellegen, 1996), and that it is this latter set-point, characteristic of the individual, that is influenced by genetic variability (although the range and/or frequency of variation of any trait may itself be a heritable trait). More specifically, trait set-points are determined by genetic, epigenetic (Wong et al., 2005), and lasting environmental influences. The genetic component may include those environmental influences that were genetically induced and which therefore contribute to the trait's broad heritability. Some life experiences that have lasting effects on trait set-points, especially during one's youth, are not genetically induced and therefore, together with most of the epigenetic influences, constitute the nonheritable component of the set-point. But many relevant life experiences, especially in adulthood, produce only transitory (although sometimes large) changes in trait intensity, above or below the set-point value. Therefore, MZ correlations based on single measurements almost always underestimate the heritability of the trait's stable value.

All test scores, of course, are subject to *measurement errors* due to carelessness, misunderstandings, inattention, and the like. Repeating the test or questionnaire tomorrow is unlikely to yield the identical score. But if Jones got a standard score of 60 on the MPQ Well Being scale last year and only 40 this year, the 2-sigma difference is not likely to be measurement error. Jones is truly less happy now than he was a year ago, either because he lost his job or his wife or his freedom in the interim (or because he had gotten a job or a wife or his freedom just before he took the test last year). If we repeat Jones's MPQ right away this year and average the two scores, we can minimize measurement error but that average may still be well below his stable or set-point level.

Received 17 August, 2006; accepted 25 August, 2006.

†*David Lykken passed away shortly after this paper was accepted for publication.*

Address for correspondence: William G. Iacono, Department of Psychology, 75 East River Road, Minneapolis, MN 55455 USA. E-mail: wiacono@tfs.psych.umn.edu

Suppose that, instead of two measurements 10 years apart, we had obtained 10 trait measurements just one year apart. It seems plausible that the mean of those 10 scores would be largely free of the transient variations that contributed to those low retest correlations. We know that the correlation between pairs of MZ twins separated in infancy and reared apart (MZA twins) is itself an estimate of broad heritability. We also know (Bouchard et al., 1990) that, for most traits, R_{MZA} is about equal to the correlation for MZ twins reared together (R_{MZT}). Therefore, the R_{MZT} of those 10 mean scores would yield a good estimate of the broad heritability of most trait set-points. Obtaining many repeated measurements months or years apart is not an easy task, but there is an easy way to get a reasonable estimate of set-point heritability with just two measurements.

The similarity of each twin at Time 1, with his co-twin at Time 2, is determined by the heritable component of their set-points for the trait in question. The measure of this similarity, the cross-twin cross-time correlation (R_{CT}), is diminished by any nongenetic difference between their set-points, and especially by transitory environmental influences that cause Twin A's trait-value to vary from his set-point at Time 1 differently than the deviation of Twin B's trait-value from B's set-point at Time 2. The retest, or within-twin cross-time correlation (R_{WT}), is diminished by these same environmental influences that are uncorrelated between occasions. Therefore, a reasonable estimate of the similarity of the twins' stable trait set-points can be obtained by dividing R_{CT} by R_{WT}, computed over a sufficient time interval.

Although McGue, Bacon, and Lykken (1992) reported the R_{CT} and R_{WT} correlations for the MPQ variables, we failed to compute their ratios, which reveal the best estimate of mean heritability of the stable scale scores to be .74 instead of .51, while that for the three factors is .76 instead of .55. These data also indicate that at least 20% of the individual differences in set-point values in this sample were due to lasting effects of unshared environmental influences.

Some Illustrative Data

Retests With the Multidimensional Personality Questionnaire

The Minnesota Twin Registry (Lykken et al., 1990) was created by identifying all twins born in Minnesota from 1936 through 1955, then recruiting all locatable intact pairs, then aged 30 through 50, by mail. The MPQ was one of several questionnaires completed by more than 4000 middle-aged pairs. Two hundred and five pairs completed the MPQ again some 3 years after their first testing.

They included 131 MZ and 74 dizygotic (DZ) twin pairs. The data for the 141 female and 64 male pairs were correlated separately and then averaged. Table 1 displays the correlations for the three MPQ factors plus the mean correlation for the 13 scales. The mean of the *R*s obtained by the MZ pairs on the two occasions averaged .54 for scales and .57 for factors, but the *R*s of the means of the scores obtained by the MZs on the two occasions averaged .62 for the scales and .63 for factors, because the means of two (or more) scores provide better estimates of the stable or set-point values of the several traits. When the cross-twin cross-time correlations (R_{CT}) are disattenuated by dividing by the corresponding retest correlations — giving the best estimate of the stable heritability of these traits — then the mean of the MZ values, averaged over the 13 scales, equals .73 while the mean for the three MPQ factors is .71. These results are very similar to those found by McGue et al. (1992), cited above, with a somewhat younger sample retested after a 10-year interval. Note that the MPQ's

Table 1

Mean Correlations of Multidimensional Personality Questionnaire Scale and Factor Scores

Item	First *R*	Second *R*	Mean of *R*s	*R* of means	Retest R_{WT}	R_{CT}	R_{CT}/R_{WT}
Scale means							
MZ	.58	.51	**.54**	**.62**	.60	.44	**.73**
DZ	.22	.25	**.24**	**.24**	.52	.13	**.25**
Positive emotionality							
MZ	.63	.45	**.54**	**.59**	.62	.44	**.71**
DZ	.05	.15	**.10**	**.03**	.64	.03	**.05**
Negative emotionality							
MZ	.59	.54	**.57**	**.66**	.58	.39	**.67**
DZ	.45	.27	**.36**	**.33**	.58	.17	**.29**
Constraint							
MZ	.57	.62	**.60**	**.65**	.63	.46	**.74**
DZ	.23	.27	**.25**	**.30**	.54	.14	**.26**

Note: Obtained from 131 MZ and 74 DZ (64 male and 141 female) twin pairs, aged from 30 to 55, who took the MPQ twice, 3 years apart.

https://doi.org/10.1375/twin.10.1.168 Published online by Cambridge University Press

Positive Emotionality factor appears to be strongly emergenic (Li, 1987; Lykken, 1982, 2006; Lykken et al., 1992).

Retests With Self-Rating Items

The subsample of Registry twins also completed a set of 65 self-rating items on the two occasions, 3 years apart. The response form began as follows:

> In this next set of questions you are asked to compare yourself with other people on the specified trait or ability — where do you think you would rank compared to other people of your age and sex? This will be difficult for some of these traits but please make the best estimate that you can. For each item in this section, please use the following rating scale:

1	2	3	4	5
Lowest 5%	Lower 30%	Middle 30%	Higher 30%	Highest 5%

1. Abstract Intelligence: The ability to solve intellectual problems, to understand complicated issues, to figure things out: 'school intelligence'.

2. Contentment: Taking the good with the bad, how happy and contented are you on the average now, compared with other people?

Correlations for six of the items, plus the mean values for all 65 items, are given in Table 2.

The MZ pairs correlated .56 in self-rated Abstract Intelligence on the first occasion and .41 on the second, giving a mean R_{MZ} = .49. But the means of the two self-ratings gave an R_{MZ} of .63, showing that both self-ratings by each twin deviated randomly from that twin's average self-rating, which would yield the maximum co-twin similarity. The best estimate of that true similarity, the cross-twin cross-time correlation divided by the retest correlation, yields a disattenuated R_{MZ} of .85, about what one might expect if, instead of self-ratings, one were correlating IQ test scores! The correlations of the DZ twins increase from .22, for the mean of the two R_{DZS}, to .29, for the R_{DZ} of the means of the two self-ratings, to a disattenuated R_{DZ} of .41.

Of the six listed self-rated items, Abstract IQ, Contentment, and Social Closeness are polygenic additive traits, while the other three have very small DZ correlations and appear to be emergenic.

Retests With Interests and Talents

That same sample of middle-aged Registry twins also rated 100 Occupational Interests, 120 Leisure-Time Interests, and completed 40 Talent Self-Ratings on those two occasions 3 years apart. Table 3 provides

Table 2

Twin Pair Responses to 65 Self-Rating Items Twice, 3 Years Apart

Item	First *R*	Second *R*	Mean of *R*s	*R* of means	Retest R_{WT}	R_{CT}	R_{CT}/R_{WT}
Abstract IQ							
MZ	.56	.41	**.49**	**.63**	.61	.52	**.85**
DZ	.25	.18	**.22**	**.29**	.56	.23	**.41**
Contentment							
MZ	.38	.38	**.38**	**.48**	.44	.31	**.73**
DZ	.05	.11	**.11**	**.17**	.44	.16	**.36**
Creative							
MZ	.37	.26	**.31**	**.40**	.58	.30	**.54**
DZ	.16	–.02	**.07**	**.10**	.58	.08	**.14**
Irritable							
MZ	.24	.31	**.27**	**.36**	.33	.22	**.55**
DZ	–.07	.06	**.00**	**.06**	.33	.01	**.03**
Choleric							
MZ	.36	.28	**.32**	**.50**	.40	.41	**.89**
DZ	.07	.06	**.07**	**.06**	.40	.01	**.03**
Social closeness							
MZ	.61	.55	**.58**	**.64**	.65	.56	**.56**
DZ	.03	.20	**.11**	**.14**	.47	.12	**.26**
Means of 65 items							
MZ	.33	.32	**.33**	**.41**	.52	.31	**.58**
DZ	.11	.14	**.12**	**.16**	.52	.11	.22

Note: 102 MZ pairs (49 male and 53 female) and 96 DZ pairs (44 male and 52 female) of middle-aged (30–55) twins.

The twin correlations for six of the individual items plus the means for all 65 items are given.

https://doi.org/10.1375/twin.10.1.168 Published online by Cambridge University Press

Table 3

Twin Pair Responses to 100 Occupational Interest, 120 Leisure-Time Interest, and 40 Talent Items Twice, 3 Years Apart

Item	First *R*	Second *R*	Mean of *R*s	*R* of means	Retest R_{WT}	R_{CT}	R_{CT}/R_{WT}
Hunt and fish							
MZ	.63	.50	**.56**	**.60**	.81	.53	**.65**
DZ	.48	.36	**.42**	**.46**	.79	.42	**.53**
Novels and games							
MZ	.52	.43	**.48**	**.50**	.66	.38	**.58**
DZ	.05	.16	**.11**	**.12**	.74	.09	**.13**
120 leisure-timeinterests							
MZ	.32	.22	**.25**	**.31**	.58	.26	**.44**
DZ	.09	.18	**.14**	**.12**	.66	.07	**.10**
Husbandry							
MZ	.64	.60	**.62**	**.70**	.78	.63	**.81**
DZ	.35	.31	**.33**	**.37**	.72	.31	**.43**
Artist, crafts							
MZ	.41	.51	**.46**	**.55**	.69	.47	**.68**
DZ	.15	.13	**.14**	**.13**	.71	.08	**.11**
100 occupational interests							
MZ	.41	.47	**.44**	**.50**	.65	.38	**.58**
DZ	.30	.08	**.19**	**.19**	.68	.10	**.17**
Artistic talent							
MZ	.44	.34	**.39**	**.47**	.69	.37	**.59**
DZ	.08	−.03	**.02**	**.05**	.72	.05	**.08**
Political talent							
MZ	.23	.35	**.29**	**.35**	.63	.29	**.47**
DZ	.32	.24	**.28**	**.33**	.60	.28	**.43**
40 talents							
MZ	.48	.56	**.52**	**.56**	.78	.49	**.63**
DZ	.17	.07	**.13**	**.13**	.78	.10	**.13**

Note: 102 MZ and 96 DZ pairs of middle-aged twins.

The twin correlations for six of the individual items plus the means for each category are given.

the mean twin corrections for these three questionnaires, plus two sample items from each of the three.

Interest in recreational Hunting and Fishing displays strong and lasting effects of shared family influence; the DZ correlation is 75% of the MZ value. Both men and women have quite definite opinions about these activities, so the 3-year retest correlations are unusually high and the disattenuated estimate of MZ co-twin similarity is only 16% higher than the mean of the R_{MZ} estimates on the two occasions. For the average Leisure-Time Interest, in contrast, the MZ R_{CT}/R_{WT} value is 76% greater than the average single-shot R_{MZ}. As is true also for the averages of Occupational Interests and Talents, the average Leisure-Time Interest tends toward emergenesis. However, while Artistic Talent, and interest in being an Artist or Craftsman, are clearly emergenic, interest in Husbandry (building, repairing) and Political Talent both tend to run in families.

It is worth emphasizing that the substantial heritabilities listed in Tables 2 and 3 were not for multi-item tests or inventories but, rather, for individual self-rating items.

Retests of Heart Rate and Blood Pressure

In the longitudinal Minnesota Twin/Family Study (MTFS), begun in 1986, twins born in Minnesota either 11 or 17 years previously were identified from birth records and recruited to come to the university with their parents for a 1-day session of varied assessment (Iacono & McGue, 2002). Participants were contacted annually after this intake visit and invited to return for follow-up visits at 3-year intervals. Various measures were repeated at some of these later visits.

Heart rate (HR) and BP were assessed on three occasions during each 1-day visit and the three values averaged for reliability. Table 4 shows that systolic BP and HR gave MZ correlations averaging .50 on the two occasions, 3 years apart. When the two measurements are averaged, the MZ correlations for these

https://doi.org/10.1375/twin.10.1.168 Published online by Cambridge University Press

Table 4

Correlations of Blood Pressure and Heart Rate of 669 MZ and 433 DZ Twin Pairs

Item	First *R*	Second *R*	Mean or *R*s	*R* of means	Retest R_{WT}	R_{CT}	R_{CT}/R_{WT}
Systolic blood pressure							
MZ	.42	.58	**.50**	**.60**	.29	.28	**.94**
DZ	.39	.30	**.34**	**.39**	.36	.20	**.57**
Diastolic blood pressure							
MZ	.32	.38	**.35**	**.40**	.12	.11	**.93**
DZ	.25	.28	**.26**	**.28**	.29	.10	**.34**
Heart rate							
MZ	.44	.53	**.49**	**.56**	.34	.30	**.87**
DZ	.30	.21	**.26**	**.31**	.37	.17	**.45**

Note: Twin pairs measured twice, 605 pairs at ages 11 and 14, and 497 pairs at ages 17 and 20.

mean values are somewhat larger. Not surprisingly, the diastolic BP readings were significantly less reliable than the other two measures and the MZ correlations correspondingly smaller. The DZ correlations of the BP and HR means are significantly higher than one half of the MZ values, suggesting that sharing a day of somewhat stressful assessment increased temporarily the DZ's similarity in these cardiovascular variables. But the ratio R_{CT}/R_{WT} indicates that the BP and HR set-points — the levels to which these variables return when these young twins relax after stimulation — have heritabilities of about 90%.

Retests With the Wechsler Intelligence Scales

The claim that *all* traits vary, due to the slings and arrows and so forth, about their stable set-points might seem excessive when one thinks of traits like intelligence (although I've always known that I am dumber sometimes than at others). A brief form (composed of the Information, Vocabulary, Block Design, and Picture Arrangement subtests) of the Wechsler IQ test was administered to both the MTFS 11-year-olds (who received the Weschler Intelligence Scale for Children — Revised [WISC-R]) and the 17-year-olds (who received the Weschler Adult Intelligence Scale — Revised [WAIS-R]) on their intake visits and again on their second follow-up (when both groups received the WAIS-R) when they were aged either 17 or 24 (at age 24, only the Information and Block Design subtests were administered). The intraclass correlations for their pro rata IQ scores are shown in Table 5.

The correlations of the mean IQs were .85 for the younger MZ twins and .83 for the older sample, compared to mean single-measure correlations of .79 and .77, respectively. It will be obvious that the mean of two independent measurements, while a better estimate of the underlying stable value than any single measurement, would not be as valid as the mean of numerous measurements, also repeated far enough apart so that the transient environmental influences dissipate. Once again, the cross-twin cross-time correlation, divided by the retest correlation, provides a reasonable estimate of the correlation one might obtain of the means of repeated measurements. These values are listed for both samples in Table 5. This ratio, the disattenuated estimate of the true correlation of the MZ's set-point values, equals .92 for the younger sample, and .91 for the older twins.

Table 5

Pro Rata Wechsler IQ Correlations for MTFS Twins Tested at Ages 11 and 17, and for Another Sample Tested at Ages 17 and 24

	Age 11 *R*	Age 17 *R*	Mean of *R*s	*R* of means	Retest R_{WT}	R_{CT}	R_{CT}/R_{WT}
310 MZ pair	.76	.82	.79	.85	.77	.71	.92
167 DZ pair	.48	.51	.50	.52	.76	.42	.56
	Age 17 *R*	Age 24 *R*	Mean of *R*s	*R* of means	Retest R_{WT}	R_{CT}	R_{CT}/R_{WT}
186 MZ pair	.78	.75	.77	.83	.79	.72	.91
113 DZ pair	.41	.26	.34	.36	.76	.30	.40

Note: The data show that the retest means provide better estimates of the stable or IQ set-points than either measurement alone and that the MZ ratios of cross-twin cross-time to retest correlations provide an estimate of the true heritability of the stable IQ values.

https://doi.org/10.1375/twin.10.1.168 Published online by Cambridge University Press

Summary

All psychological and psychophysiological traits vary, often widely, about their stable set-point values, due in small part to measurement error, but due mainly to real variation in the current trait value resulting from transitory environmental influences. Because it is this stable set-point that embodies the genetically determined component of the trait, twin and family data based on one-time trait measurements *must* underestimate true trait heritability. Estimates based on the means of repeated measurements, taken far enough apart so as not to be affected by the same transitory influence, would yield accurate and substantially larger heritability estimates for HR and BP, traits of personality, interests, talents, and even IQ. Such data are rare and costly so an alternative method, requiring only two measurements sufficiently far apart, is described and illustrated.

The cross-twin cross-time correlation, R_{CT}, reflects co-twin similarity in the heritable component of their set-points, reduced by any nongenetic, environmentally produced difference between their set-points, and reduced especially by transitory, unshared environmental influences that vary each twin's trait scores above or below their set-points on each occasion. The retest, or within-twin cross-time correlation, R_{WT}, is diminished by these same uncorrelated influences so that the ratio, R_{CT}/R_{WT}, estimates the true or disattenuated similarity of the co-twins' trait set-points. For the *means* of personality scales, and even for individual self-rating, interest, and talent *items*, these disattenuated heritabilities were from 30% to 81% higher than those based on single measurements. For HR and systolic BP, broad heritabilities estimated from MZ correlations were 49% and 50%, respectively, when based on single measurements, but 87% and 94% when based on the respective R_{CT}/R_{WT} ratios. Even for pro rata Wechsler IQ scores, heritability increased from .79 to .92 when estimated in this way. These data also indicate that from 10% to 40% of individual differences in the stable set-points of these traits is due to nongenetic influences, including epigenetic factors and lasting effects of adventitious experiences.

Acknowledgment

Supported by National Institutes of Health grants DA 05147 and AA 09367.

References

Bouchard, T. J., Lykken, D. T., McGue, M. M., Segal, N., & Tellegen, A. (1990). The sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science, 250*, 223–228.

Carey, G. (2003). *Human genetics for the social sciences*. Thousand Oaks, CA: Sage.

Iacono, W. G., & McGue, M. (2002). Minnesota Twin Family Study. *Twin Research, 5*, 482–487.

Li, C. C. (1987). A genetical model for emergenesis. *American Journal of Human Genetics, 41*, 517–523.

Lykken, D. T. (1982). Research with twins: The concept of emergenesis. *Psychophysiology, 19*, 361–373.

Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Chiccetti & W. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 3–39). Minneapolis, MN: University of Minnesota Press.

Lykken, D. T. (2006). The mechanism of emergenesis. *Genes, Brain and Behavior, 5*, 306–310.

Lykken, D. T., Bouchard, T. J., McGue, M., & Tellegen, A. (1990). The Minnesota Twin Registry: Some initial findings. *Acta Geneticae Medicae et Gemmellologiae, 39*, 35–70.

Lykken, D. T., Bouchard, T .J., Jr., McGue, M., & Tellegen, A. (1992). Emergenesis: Genetic traits that may not fun in families. *American Psychologist, 47*, 1565–1577.

Lykken, D. T., & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science, 7*, 186–189.

McGue, M., Bacon, S., & Lykken, D. T. (1992). Personality stability and change in early adulthood: A behavioral genetic analysis. *Developmental Psychology, 29*, 96–109.

Tellegen, A., & Waller, N. G. (1994). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In S. R. Briggs & J. M. Cheek (Eds.), *Personality measures: Development and evaluation* (Vol. 1). Greenwich, CT: JAI Press.

Wong, A. H. C., Gottesman, I. I., & Petronis, A. (2005). Phenotypic differences in genetically identical organisms: The epigenetic perspective. *Human Molecular Genetics, 14*, 11–18.

https://doi.org/10.1375/twin.10.1.168 Published online by Cambridge University Press