Twin Research (2000) 3, 224–233

www.nature.com/tr

Parent ratings of temperament in twins: explaining the 'too low' DZ correlations

Kimberly J Saudino[1], Stacey S Cherny[2] and Robert Plomin[3]

[1]Psychology Department, Boston University, USA
[2]Wellcome Trust Centre for Human Genetics, University of Oxford, UK
[3]Institute of Psychiatry, University of London, UK

Twin studies of child temperament using objective measures consistently suggest moderate heritability for most dimensions. However, parent rating measures produce unusual patterns of results. Intraclass correlations for identical (MZ) twins are typically high, whereas fraternal (DZ) twin intraclass correlations are much lower than would be predicted from an additive genetic model. The 'too low' DZ correlations can be explained by parent-rating biases that either exaggerate the differences between DZ twins (contrast effects) or that inflate the similarity of MZ twins (assimilation effects), or by the presence of non-additive genetic variance. To evaluate the three possible explanations, we used model-fitting procedures applied to parent-rating data averaged across 14, 20, 24, and 36 months of age in a sample of 196 twin pairs participating in the MacArthur Longitudinal Twin Study. The data were best described by a model that included contrast effects. Implications for non-twin research are discussed. Twin Research (2000) 3, 224–233.

Keywords: twins, temperament, genetics, rater bias, model fitting

Introduction

Research employing the twin design provides the strongest evidence of genetic influences on individual differences in temperament during infancy and early childhood. However, twin studies assessing temperament via parent-rating measures, the most common method employed to assess temperament, frequently produce an unusual pattern of results. With such measures, identical (monozygotic, MZ) co-twin resemblance for temperament dimensions, as indexed by intraclass correlations, is typically moderate; whereas fraternal (dizygotic, DZ) co-twin resemblance is very low, often near zero or even slightly negative.[1–3] This puzzling outcome is particularly evident in, but not exclusive to, rating measures that require parents to make global judgements of their child's behaviour.[4] Because MZ twins are more similar than DZ twins for parent-report measures of temperament, these results provide evidence of genetic influence. Nonetheless, the low DZ twin resemblance is puzzling because the simple (additive) genetic model predicts that DZ twin similarity should be at least half that of the MZ twins due to the fact that DZ twins are, on average, 50% similar genetically, whereas MZ twins are genetically identical. The pattern of very low DZ correlations is significant because it implies that DZ twins are perceived as hardly any more similar than two randomly paired children.

Correspondence: Kimberly J Saudino, Psychology Department, Boston University, 64 Cummington St, Boston, MA 02215, USA. Tel: 617 353 3679; Fax: 617 353 6933; E-mail: ksaudino@bu.edu
Received 10 April 2000; accepted 18 April 2000

Possible explanations for low DZ resemblance

The problem of 'too low' DZ correlations, as it has come to be described in the infant and child temperament literature, has been explained by assimilation effects, contrast effects, or by the presence of non-additive genetic variance. Both assimilation effects and contrast effects refer to parental rating biases that artificially increase differences between MZ and DZ correlations and, therefore, result in overestimates of genetic influence. Assimilation effects refer to rater biases that accentuate similarities between MZ twins. Perhaps as a result of the high physical similarity between MZ twins, parents may tend to overestimate the degree of temperamental similarity. The issue here, then, is not that the DZ correlations are too low, but that the MZ correlations are too high.

Contrast effects arise when parents rating the temperament of co-twins make comparisons that magnify existing behavioural differences. The greater the actual behavioural difference between co-twins, the greater the tendency to exaggerate the differences between co-twins. Thus, for genetically influenced traits, such as temperament, MZ twins, who are more behaviourally alike, would be less prone to rater contrasts.[5] Because contrast effects operate more strongly for DZ twins, their correlations will be 'too low' as compared with MZ correlations.

The third possible explanation, non-additive genetic variance, refers to effects of genes that are not linear and additive. For example, if there is dominance among alleles (alternate forms of a gene), or if a trait is influenced by epistasis (interaction of alleles across loci), the phenotypic expression of the trait does not represent the sum of the average effects of alleles. MZ twins share all non-additive genetic effects, whereas DZ twins and other first-degree siblings share only a quarter of genetic variance due to dominance and even less variance due to epistasis. Thus, if non-additive genetic variance contributes to a trait, DZ twin similarity will be less than one half that of MZ twins.

Evaluating the alternative explanations: objective measures of temperament

The use of objective measures of temperament provides one way to evaluate the possible explanations for 'too low' DZ correlations. If the unusual outcomes from parent-rating studies of temperament are due to parental rating biases, then objective measures should show a more reasonable pattern of results. Specifically, if parent ratings are biased by assimilation effects, then objective measures of temperament should show substantially lower MZ correlations. Similarly, if contrast effects are operating for parent ratings, then objective measures of temperament should show higher DZ correlations. On the other hand, if the low DZ similarity is due to non-additive genetic variance, then objective measures and parent ratings would be expected to show a similar pattern of results (ie DZ correlations less than one-half MZ correlations for both measures).

Although the bulk of twin research in infancy and early childhood has relied on parental ratings of temperament, there is a handful of studies that have employed more objective measures of temperament such as tester and observer ratings or mechanical measures.[2,6–11] When temperament is assessed via these more objective measures, DZ similarity is not inappropriately low. A good example comes from the MacArthur Longitudinal Twin Study (MALTS), a collaborative longitudinal study of twins that focuses on individual differences in temperament, emotion, and cognition from infancy to early childhood.[12] MALTS provides a unique opportunity to evaluate the issue of possible parental biases because it includes many observed behavioural measures of temperament in addition to parent ratings.

Figure 1 presents intraclass correlations for parent and observer-rated temperament scores averaged across 14, 20, 24, and 36 months of age in a sample of approximately 200 pairs of twins in MALTS. These results are summarised from two previous longitudinal analyses separately examining sources of continuity and change in parent and observer-rated temperament.[13,14] Parent ratings of emotionality, activity, shyness, and attention/persistence were obtained using the Colorado Childhood Temperament Inventory[15] (CCTI). The Infant Behavior Record[16] (IBR) provided observational measures of behaviours conceptually related to emotionality, activity, and attention/persistence (ie affect/extraversion, activity, and task orientation, respectively) based on previous factor analyses of the IBR items.[7] A measure of observed shyness was obtained from behavioural observations of each child's initial reaction to the entrance of two female examiners to the home. As can be seen in Figure 1, parent ratings of temperament produced a pattern of moderate MZ correlations and negative DZ twin correlations. In fact, the DZ correlations for activity and attention/persistence were significant ($P < 0.05$), indicating that for these dimensions, DZ co-twins are perceived as having opposing behavioural tendencies.

Observational measures of similar temperament dimensions tell a different story. When temperament was assessed by observer ratings, DZ correlations for all dimensions were positive and significantly different from zero ($P < 0.05$). The pattern of MZ–DZ correlations was consistent with additive genetic expectations with DZ twin resemblance approximately half that of the MZ twins. Moreover, MZ correlations are similar for both parent ratings and observer ratings, which argues against the assimilation hypothesis. Overall, the evidence from MALTS points toward parental contrast effects.

Evaluating the alternative explanations: model-fitting analyses

However impressive, data comparing twin correlations for parent ratings and observational measures of temperament provide only an indirect test of contrast effects. The problem with this approach is that the two methods of assessing temperament may be tapping different behavioural tendencies such that comparing the results from the two methods may be akin to comparing apples and oranges because parent and observer ratings assess children in different contexts. Parent ratings provide valuable information about a child's usual behaviour across many day-to-day situations. In comparison, behavioural observations are typically based on brief behavioural samples that involve assessing the child's reactions to mildly stressful, standardised situations, usually within a laboratory environment. Both methods have advantages and disadvantages, but the key issue is the extent to which they assess the same behaviour. Many studies have found that parent ratings and observational measures of temperament are only weakly correlated.[14,17–19] Thus, it

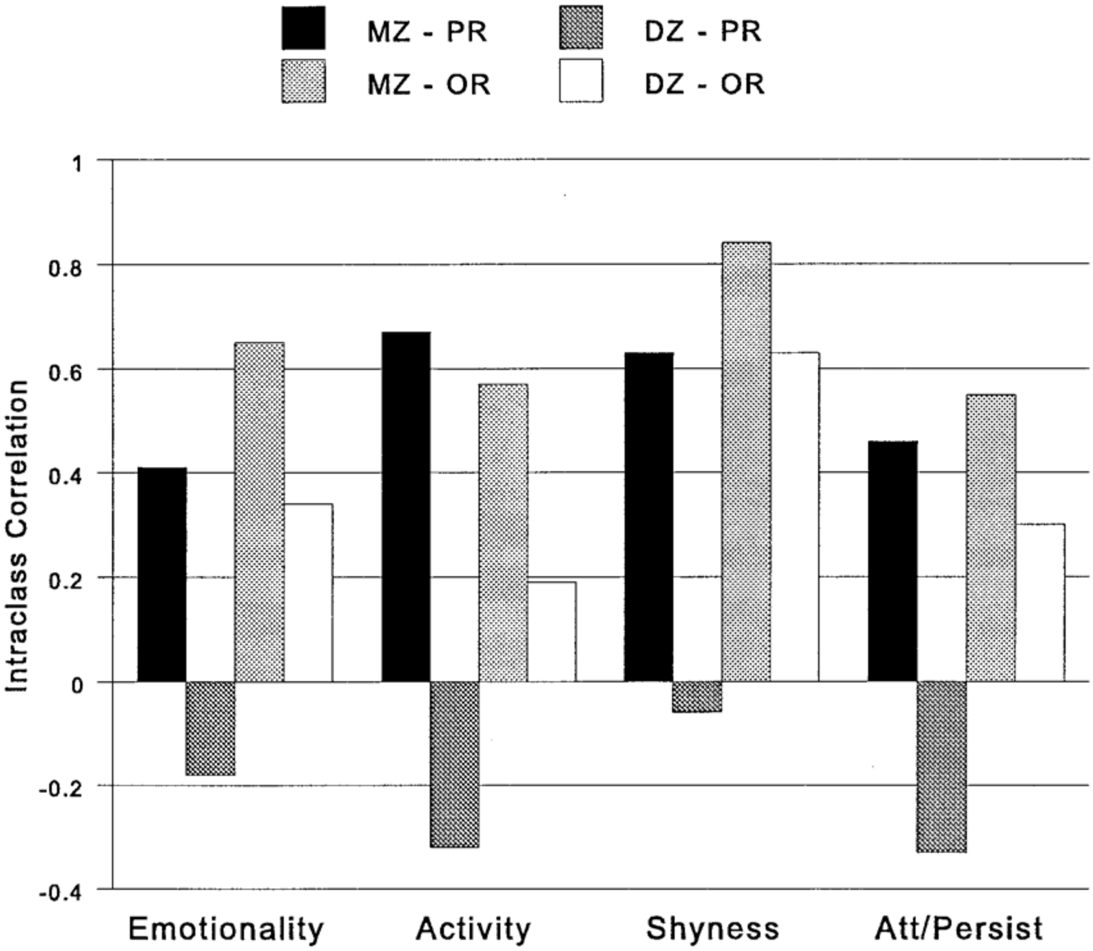

Figure 1 MZ and DZ twin intraclass correlations for parent-rated (PR) and observer-rated (OR) temperament scores averaged across 14, 20, 24, and 36 months of age from the MacArthur Longitudinal Twin Study

appears that the two methods yield different information regarding a child's temperament. It is possible, therefore, that contextual differences between parent and observer ratings of temperament may be responsible for the different outcomes when these measures are used in twin research.

By applying model-fitting techniques to parent-rating data it is possible to obtain a more direct evaluation of the three alternative explanations for the 'too low' DZ correlations. A strength of model-fitting approaches lies in the ability to evaluate alternative models and determine which model provides the best description of the data. Previous research comparing additive and non-additive models fit to parent ratings of temperament in twins found that a non-additive model fit best.[20] On the basis of this finding, the authors concluded that contrast effects were operating. However, what they tested for was non-additive genetic variance, not contrast effects. A more explicit test of contrast effects is possible. In the present paper, we fit three models to the data:

(a) the standard additive genetic model;

(b) a model that includes nonadditive genetic influence; and

(c) a sibling interaction model that tests for the presence of assimilation or contrast effects.[21,22]

Given the negative DZ correlations found previously in MALTS, we expected that the additive model would provide the poorest fit to the data. We also expected that the fit of the non-additive model would be somewhat better than that of the additive model, but that the sibling interaction model would indicate the presence of contrast effects and would provide the best overall fit to the data.

Methods

Sample

The MALTS sample was recruited from monthly reports of births from the Colorado Department of Health. Twins were selected preferentially for higher birth weight (>1700 g) and gestational age (>34 weeks), although some healthy lower weight infants were included in the sample (4% weighed less than 1700 g). The average birth weight for the sample was 2579 g (SD = 469). (Reznick et al[23] give more details regarding the MALTS sample). The present analyses include 196 same-sex pairs of twins (101 MZ, 95 DZ) who had parent-rated temperament data across the ages of 14, 20, 24, and 36 months (ie all twin pairs who had data at all four time points). Twin zygosity was established using physical similarity criteria. Within each zygosity group there were approximately equal numbers of male and female twin pairs. (See Plomin et al[2] for zygosity diagnosis procedures.)

Measures

At each age, both parents rated the temperament of their twins on the Colorado Childhood Temperament Inventory[15] (CCTI) which had been modified to include separate and distinct scales of Shyness and Sociability.[5] The revised measure contains general statements describing the temperament dimensions of Emotionality, Activity, Sociability, Shyness, and Persistence, for example 'Child cries easily' or 'Child is very energetic'. Parents were asked to rate each statement on a 5-point Likert scale ranging from 1 (strongly disagree; not at all like the child) to 5 (strongly agree; a lot like the child). Because aggregating across raters increases the reliability of a measure by reducing the error variance associated with a single rater,[24] mid-parent scores were created by averaging across mothers and fathers. Although we present the results for the analyses of mid-parent scores, it should be noted that analyses conducted separately for mothers' and fathers' ratings of temperament yielded the same pattern of results.

Internal consistency reliabilities in the present study for the five CCTI scales ranged from 0.73 to 0.89. In MALTS, age-to-age stability can be used as a lower limit estimate of reliability for the parent ratings. Even from 14 to 36 months, the largest age interval, significant stability was found (0.40, 0.52, 0.39, 0.42, and 0.42, for the five CTTI traits, respectively). Stability across adjoining age intervals ranged from 0.49 to 0.69. Considering that there is much developmental change across the transition from infancy to early childhood, these stabilities are impressive.

Model-fitting

Each of the three alternative models was fit to twin variance/covariance matrices using Mx[25] maximum-likelihood model-fitting procedures. The three models are depicted as path diagrams in Figure 2. For each model, the observed phenotypic variances of each twin are represented by the two rectangles. The circles represent latent genetic and environmental variables. The curved double-headed arrows indicate correlations between the variables they connect. The single-headed arrows represent paths (a, c, d, and e), partial regressions of the measured variable on the latent factor.

Additive model According to the additive model (Figure 2a), phenotypic variation is assumed to be due to three latent variables: additive genetic effects (A), shared environmental effects (C), and nonshared environmental effects (E). Additive genetic influences refer to the sum of the average effect of all genes that influence a trait. Identical (MZ) twins are genetically identical, having 100% of their genes in common, whereas fraternal (DZ) twins share on average 50% of their segregating genes. Therefore, as indicated in the path diagram, MZ co-twins are correlated 1.0 and DZ co-twins are correlated 0.50 for additive genetic effects. Shared environment refers to environmental influences that are shared among co-twins (eg shared parental behaviour, attending the same school) making them similar. Because in MALTS, co-twins live together, both MZ and DZ twins are correlated 1.0 for shared environment. Finally, non-shared environmental influences are those environmental factors that are unique to each member of a twin pair and that make twins different from each other (eg illnesses or accidents, measurement error) and are therefore, depicted in the path diagram as residual arrows for each twin representing the remaining variance not explained by genes or shared environment. The single-headed arrows a, c, and e represent paths, partial regressions of the measured variable on the latent variable; a is the genetic parameter, c is the shared environmental parameter, and e is the non-shared environmental parameter.

Non-additive model With the simple twin design, shared environmental effects and non-additive genetic effects are confounded and cannot be estimated within the same model because both are based on the extent to which the DZ covariance deviates from one-half the MZ covariance (ie shared environmental effects: DZ $r > \frac{1}{2}$MZ r; non-additive genetic effects: DZ $r < \frac{1}{2}$MZ r). Therefore, the non-additive model (Figure 2b) substitutes genetic dominance (D)

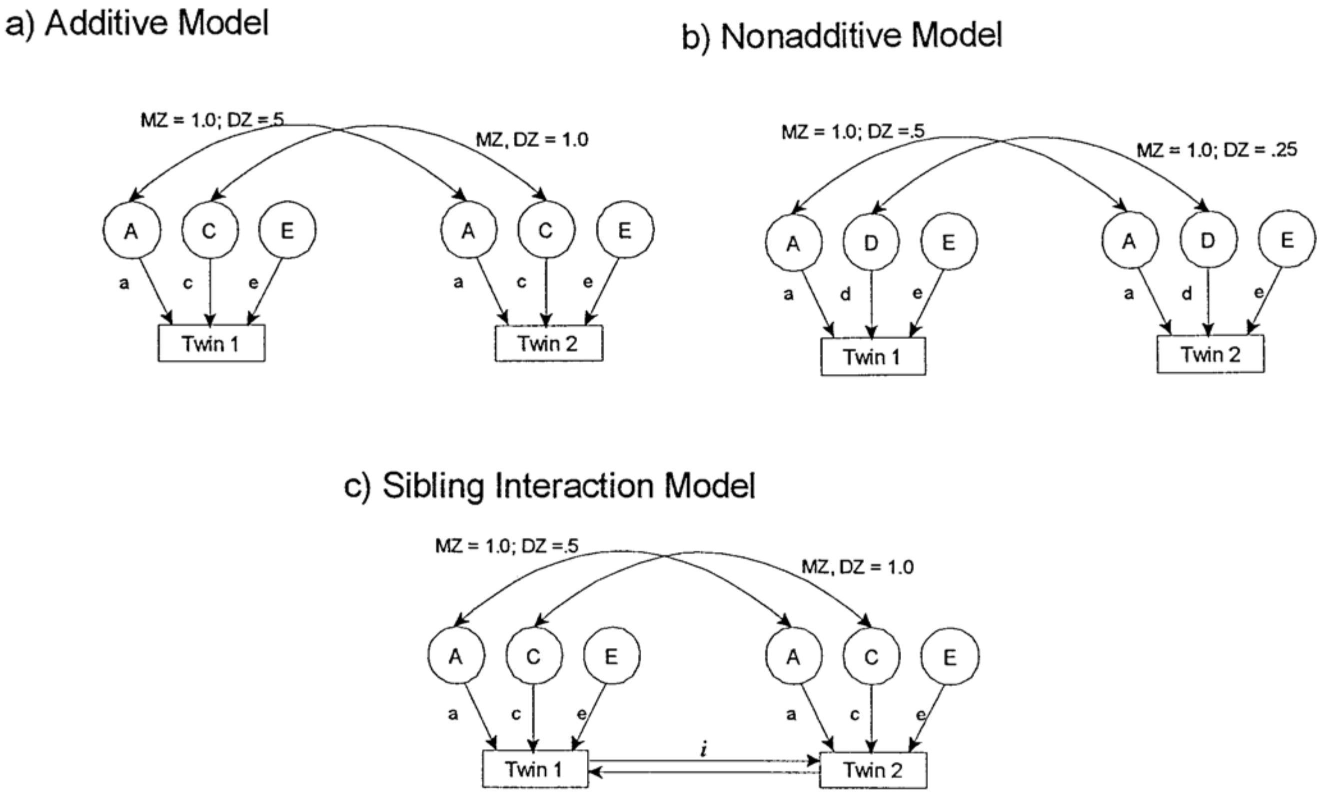

Figure 2 a Additive genetic model. Twin 1 and Twin 2 are measured variables for the two twins (ie the phenotype of each twin). A, C, and E are latent variables representing additive genetic variance, shared environmental variance, and non-shared environmental variance, respectively. The curved two-headed arrows indicate correlations between the variables they connect. The single-headed arrows a, c, and e represent paths, partial regressions of the measured variable on the latent variable. a is the genetic parameter, c is the shared environmental parameter, and e is the non-shared environmental parameter b Non-additive genetic model. D represents non-additive genetic variance. d is the nonadditive genetic parameter (ie the partial regression of the measured variable on D) c Sibling interaction model. Interaction paths (i) from each twin's phenotype to the phenotype of their co-twin indicate that each twin's phenotype is also a function of the phenotype of their co-twin

for shared environmental effects. MZ twins share all non-additive genetic effects, whereas DZ twins share only a quarter of genetic variance due to dominance.[26] Thus, as indicated in the path diagram, MZ co-twins are correlated 1.0 and DZ co-twins are correlated 0.25 for non-additive genetic effects.

Sibling interaction model The sibling interaction model[21,22] was fit to evaluate possible assimilation and contrast effects. The critical feature of the sibling interaction model is that it takes into account variance differences between the MZ and DZ twins. In the presence of assimilation or contrast effects, the phenotypic variance of a trait will differ across twin types. Assimilation effects result in increased variance for MZ twins.[22,27] Under contrast effects, the variance of DZ twins is greater than MZ twins.[22,27] (See Eaves[27] for a discussion of sibling effects on variances.) As depicted in Figure 2c, the sibling interaction model is a modification of the additive model to include paths (i) from each twin's phenotype to the phenotype of their co-twin. Under this model each twin's phenotype is a function of additive genetic effects (A), shared environmental influences (C), non-shared environmental influences (E), and the phenotype of their sibling. That is, we are testing whether the parents' rating of one twin's temperament is influenced by their rating of the temperament of the other twin. Assimilation effects are indicated when the estimate of the interaction parameter (i) has a positive value; contrast effects ae indicated when i is negative.

The fit of each model is denoted by the χ^2 statistic. A non-significant χ^2 indicates that the model provides an adequate fit to the data. To compare the relative fit of alternative models Akaike's Information Criterion (AIC) was computed (see Neale and Cardon[22] for formula). The model with the lowest AIC was then judged to be the best-fitting model.

Results

Intraclass correlations

Twin intraclass correlations for parent ratings of temperament on the CCTI yield a pattern of DZ correlations that are much less than one-half those of MZ twins (see Table 1). Across age and dimension, parent ratings produced a pattern of moderate and statistically significant MZ intraclass correlations and near zero or negative DZ twin intraclass correlations. Many of the negative DZ correlations are significant suggesting that parents perceive their DZ twins as not simply unalike, but as having opposing temperaments. The finding of significant negative correlations strongly suggests contrast effects because neither assimilation nor non-additive genetic variance would result in such an outcome.

No clear age trends emerge in Table 1; therefore, we averaged scores across age and computed intraclass correlations for the composite scores. In addition to summarising the data across age, composite scores have the added advantage of increasing the reliability and stability of the parent ratings.[24] As can be seen in the final row of Table 1, twin intraclass correlations for the composite scores are very similar to those reported above.

MZ and DZ variances

Because of their relevance to the sibling interaction model, twin variances are presented in Table 2. DZ variances are greater than MZ variances for all variables except persistence at 14 months. In most cases, the variance differences between MZ and DZ twins were significant. These variance differences provide further evidence that contrast effects may be operating.

Comparison of alternative models

The three alternative models were fitted to the composite temperament scores for each dimension as a way of summarising the model-fitting results. Because the correlations in Table 1 and variances in Table 2 suggested no clear age trends, conducting 60 model-fitting analyses (ie three models for each dimension at each age) seems unwarranted. Table 3 presents the parameter estimates and fit statistics derived from fitting each model to the composite scores for the five CCTI temperament dimensions. For all five dimensions, the chi-squares for both the additive model and the non-additive model are significant, indicating that neither of these two models fit the data. It should be noted, however, that

Table 1 Twin intraclass correlations for parent ratings of temperament on the CCTI

	Emotionality		Activity		Sociability		Shyness		Persistence	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
14 months	0.23[b]	–0.18[a]	0.53[b]	–0.26[b]	0.38[b]	0.00	0.35[b]	–0.28[b]	0.47[b]	0.00
20 months	0.47[b]	0.00	0.51[b]	–0.23[b]	0.48[b]	0.10	0.44[b]	–0.20	0.50[b]	–0.29[b]
24 months	0.28[b]	–0.10	0.55[b]	–0.22[b]	0.42[b]	0.00	0.49[b]	0.10	0.22[b]	–0.23[b]
36 months	0.30[b]	–0.18[a]	0.56[b]	–0.32[b]	0.38[b]	–0.17[a]	0.47[b]	–0.10	0.32[b]	–0.39[b]
Composite	0.41[b]	–0.18[a]	0.67[b]	–0.32[b]	0.50[b]	0.00	0.63[b]	0.00	0.46[b]	–0.33[b]

n = 101 MZ twin pairs and 95 DZ twin pairs; [a]$P < 0.10$; [b]$P < 0.05$.

Table 2 Twin variances (and means) for parent ratings of temperament on the CCTI

	Emotionality		Activity		Sociability		Shyness		Persistence	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
14 months	10.47	12.96	6.50	14.44[b]	6.40	8.64[b]	12.74	15.58[b]	9.49	7.95
	(14.42)	(15.04)	(20.74)	(20.10)	(18.26)	(17.91)	(12.09)	(12.58)	(15.64)	(15.85)
20 months	8.83	15.24[b]	6.97	13.47[b]	7.34	10.43[b]	12.29	20.07[b]	8.07	11.70[b]
	(14.94)	(15.11)	(21.31)	(20.40)	(18.47)	(18.10)	(12.95)	(12.74)	(16.10)	(15.99)
24 months	9.06	16.48[b]	7.84	11.83[b]	7.90	10.89[b]	14.36	17.47	9.42	13.84[b]
	(14.73)	(14.84)	(21.44)	(20.37)	(18.75)	(18.30)	(12.74)	(12.86)	(16.24)	(16.25)
36 months	13.40	17.56[a]	9.12	13.76[b]	9.49	12.89[b]	18.92	20.16[b]	10.40	14.59[b]
	(14.71)	(15.01)	(20.77)	(20.19)	(18.44)	(18.16)	(13.32)	(12.70)	(16.31)	(16.92)
Composite	5.81	10.43[b]	5.06	10.24[b]	4.71	6.60[b]	7.80	12.74[b]	5.86	7.84[b]
	(14.68)	(14.98)	(21.07)	(20.22)	(18.47)	(18.07)	(12.75)	(12.76)	(16.07)	(16.25)

n = 101 MZ twin pairs and 95 DZ twin pairs; [a]variance difference between twin groups significant at $P < 0.10$; [b]variance difference between twin groups significant at $P < 0.05$.

as indicated by a lower AIC the non-additive model does provide a relatively better fit to the data than the simple additive model. Nonetheless, a non-additive model does not adequately describe the data.

The sibling interaction model is the only model to provide an adequate fit to the parent-rating data. For each temperament dimension, the χ^2 for the sibling interaction model was non-significant. Moreover, the sibling interaction model yielded the lowest AIC, indicating that this model was the better fitting model. As expected from the pattern of intraclass correlations, the negative sibling interaction parameters (i in Table 3) suggests the presence of contrast effects operating across all CCTI dimensions. The significance of the sibling interaction parameter can be evaluated by the difference in χ^2 that results when the interaction parameter is dropped from the model (ie the χ^2 difference between the sibling interaction model and the additive genetic model). For all dimensions, the sibling interaction parameter could not be eliminated from the model without a significant decrement in fit. Although not presented, analyses of mother and father ratings yielded similar results. That is, for all dimensions, the sibling interaction model with a negative interaction parameter provided the best fit to the data for both mothers' and fathers' ratings of temperament. Because it is possible that both sibling contrast effects and non-additive genetic variance could be operating, we also examined sibling interaction models that included a non-additive genetic variance parameter instead of shared environmental variance (ie ADE + i). In no case was the non-additive genetic parameter significant. Thus, once the sibling interaction term is included in the model, a non-additive genetic parameter is not necessary to explain the data.

Discussion

Although anecdotal comparisons of parent ratings and observational measures have suggested that parent ratings are prone to contrast effects, such comparisons can be criticised for comparing two different facets of temperament (ie the apples and oranges problem). Our model-fitting analyses of parent ratings of temperament provides a more direct evaluation of the three explanations for the low DZ correlations that typically emerge in twin studies employing parent ratings of temperament. As predicted, contrast effects provide the best explanation for the very low DZ correlations that occur when parents rate the temperaments of their twins. The superior fit of the sibling interaction model for all temperament dimensions suggests that parents' ratings of their twins' temperaments are not independent of each other. That is, the parents' perception of one twin's temperament is influenced by their

Table 3 Parameter estimates and fit statistics for alternative models describing parent ratings of temperament in twin siblings

	a	d	c	e	i	χ^2(df)	P	AIC
Emotionality								
Additive model	1.45	–	0.00	2.47	–	34.48 (3)	0.00	28.48
Non-additive model	0.00	1.99	–	2.10	–	26.45 (3)	0.00	20.45
Sibling interaction model	2.80	–	–[j]	1.23	–0.34	3.16 (3)	0.37	–2.84
Activity								
Additive model	2.29	–	0.00	1.80	–	81.72 (3)	0.00	75.72
Non-additive model	0.00	2.58	–	1.40	–	49.53 (3)	0.00	43.53
Sibling interaction model	5.43	–	3.25	1.45	–2.12	1.50 (2)	0.47	–2.50
Sociability								
Additive model	1.63	–	0.00	1.76	–	18.17 (3)	0.00	12.17
Non-additive model	0.00	1.77	–	1.63	–	11.76 (3)	0.00	5.76
Sibling interaction model	2.82	–	4.71	1.50	–1.97	4.73 (2)	0.09	0.73
Shyness								
Additive model	2.66	–	0.00	2.18	–	22.85 (3)	0.00	16.85
Non-additive model	0.00	2.79	–	1.98	–	10.52 (3)	0.00	4.52
Sibling interaction model	3.36	–	0.00	1.43	–0.25	2.07 (2)	0.36	–1.93
Attention/persistence								
Additive model	1.29	–	0.00	2.28	–	33.90 (3)	0.00	27.90
Non-additive model	0.00	1.75	–	1.98	–	24.83 (3)	0.00	18.83
Sibling interaction model	7.29	–	0.00	3.36	–2.90	1.35 (2)	0.51	–2.65

a: additive genetic parameter, d: non-additive genetic parameter, c: shared environment parameter, e: non-shared environment parameter, i: sibling interaction parameter. AIC = Akaike's Information Criterion. Dashes indicate parameters that are not included in the model.
[j]Parameter was not identified in the model.

perception of the temperament of the other twin. Moreover, the negative interaction parameter indicates that when rating the temperaments of their twin children, parents contrast one twin with the other and consequently magnify existing behavioural differences. Presumably DZ co-twins are more prone to contrast effects because, as compared with MZ co-twins, the actual behavioural differences between DZ co-twins is greater.

Previous research has found evidence of rater bias with parent measures of temperament. For example, in a twin study of parent EASI ratings, Neale and Stevenson[1] found that all EASI temperament dimensions displayed significant rater bias effects. However, in this previous study rater bias differs from contrast biases in that it referred to parental rating behaviour that was consistent across co-twins. Such biases might be the result of the parental personality or response style. Under this form of bias, the rater tends to consistently overestimate or underestimate the behaviour of both co-twins. This consistency across co-twins would act to inflate the similarity of both MZ and DZ twins. Although Neale and Stevenson found evidence of significant rater bias effects, the rater bias model did not adequately fit the data for three of the four temperament dimensions examined. The authors concluded that the many low and negative correlations for DZ twins may be the result of parental contrast effects, but this assumption was not tested.

The presence of contrast biases for parental ratings of temperament has important implications for behavioural genetic research examining genetic and environmental contributions to individual differences in temperament. Twin studies constitute the bulk of research exploring genetic influences on temperament, and most have relied on maternal ratings of behaviour thought to reflect temperamental dimensions.[3,15,28–31] However, contrast effects artificially increase differences between MZ and DZ correlations and thus can result in overestimates of genetic influence in twin studies. Therefore, twin studies employing parental ratings may overestimate heritability. Similarly, because contrast effects reduce the similarity of both non-adoptive and adoptive siblings, adoption studies may underestimate heritability. Together, these possibilities suggest that when sibling interaction terms are not included in models testing for genetic influences on temperament, parent-rating measures may be inadequate for precise estimates of heritability.

The greater similarity of MZ than DZ twins is nonetheless consistent with the general hypothesis of a genetic influence. Moreover, the presence of contrast effects does not vitiate behavioural genetic studies provided that the appropriate model (ie a sibling interaction model) is fit to the data. For example, in the present study, the fact that the sibling interaction model fits the data significantly better than the simple additive model allows us to conclude that genetic variance is significant because detectable interaction effects can only be found in the presence of genetic variance. Overall, our results suggest that behavioural genetic researchers should consider fitting sibling interaction models to test for the presence of contrast effects whenever parent ratings are used to assess the construct of interest.

The finding that parent-rating measures of temperament are prone to contrast effects has important implications more generally for temperament research. Parent rating scales have been the major method of assessing infant and child temperament because of their convenience, extensive behavioural sampling, ecological validity, and sound psychometric properties. Recently, however, the usefulness of these measures has been called into question as more and more research suggests that parent ratings include a subjective as well as objective component.[19,32] Parental characteristics such as SES, race, personality, and mental health can bias perceptions of temperament.[33–35] Parent expectations may also influence parental ratings of temperament.[36–39] Although the present paper examines parental expectation biases within the context of the twin design, we suggest that contrast effects are not limited to parents' ratings of twins. That is, it is reasonable to assume that whenever parents rate the temperaments of their children, they evaluate each child in the context of other children that they know well – most likely other children within the family. Thus, just as with twins, the rating of one sibling's temperament is likely to be influenced by the perceived temperament of another sibling. Because most temperament research involves the study of only one child per family, contrast effects will not be evident even when they do exist.

Evidence for contrast effects in non-twin samples comes from the Colorado Adoption Project (CAP).[40–42] In CAP, parent ratings of temperament in infancy, middle childhood, and early adolescence, consistently yield a pattern of near zero or negative correlations for both non-adoptive and adoptive sibling pairs.[43–45] These results are not, however, limited to behavioural genetic research designs. For example, Schachter and colleagues have observed that family members tend to describe siblings as different or contrasting and refer to this as 'sibling deidentification'.[46,47] Consistent with the findings of twin and adoption studies, in families with two or three children, siblings displayed significant negative correlations for global maternal ratings of easy/difficult temperament.[48]

Rooted in psychoanalytic theory, sibling deidentification theory posits that siblings seek to develop different identities as a way of coping with sibling rivalry.[46,49] Deidentification is stronger for same-sex siblings because they are more likely to compete with each other as a result of similar interests and attributes. Therefore, according to sibling deidentification theory, the behavioural differences between siblings are real and parent ratings merely reflect this. However, sibling differences resulting from the quest for one's own identity does not seem to be a plausible mechanism for explaining negative correlations between infant twins who are only just beginning to develop a sense of self. Moreover, sibling deidentification theory predicts that the more similar siblings are, the more likely that they will deidentify. However, in twin studies, DZ twins (who are less similar) show deidentification, whereas MZ twins do not. Similarly, in adoption studies, unrelated siblings show greater contrasts than related siblings. Therefore, we define contrast effects in terms of rater bias rather than sibling deidentification or competition.

Why would parents contrast the temperaments of their children? Perhaps labelling siblings' temperaments as 'X' and 'not X' based on subtle behavioural differences provides parents with a heuristic for understanding and interacting with their different children. One sibling's temperament may serve as an anchor for evaluating the temperament of other siblings. Alternatively, contrast effects might be due to the fact that parents value and seek to promote the development of the individuality of each child within a family. Carey[21] suggests that sibling constellation variables such as sex, age, birth order, or spacing might affect the contrast process. Schachter's work on sibling deidentification provides some support for this: first–second pairings of siblings showed more differences than first–third or second–third pairings, and same-sex siblings differ more than opposite-sex siblings.[47,48] Beyond this, little is known about contrast effects. It is clear, however, that only by studying more than one child per family will we be able to learn more about how and why contrast effects operate and the consequences that these parental perceptions may have on children's behaviour and self perception.

Despite our finding of contrast effects we believe that parent ratings of child behaviour are a valuable tool. There is little doubt that parents are a rich source of information about their children's typical behaviour across many situations. Such information may be lost when relying solely on observational or mechanical measures of temperament. For example, observational measures are typically brief samples of behaviour within a specific context (ie test situation) and therefore may not provide a complete picture of the child's behaviour. Similarly, mechanical measures, such as motion recorders, do not yield information about the nature and quality of temperamental behaviours (eg in what situations is the child most active?). Although they may be subjectively biased, we believe that at their core parent ratings provide some truth about their child's temperament. We would suggest, however, that our finding of contrast effects highlights the need to know more about the factors that affect the parent-rating process. Until then, research using more objective methods is needed to support the great deal that has been learned from parent-report data.

Acknowledgements

This research was supported in part by NIH grants MH-43899, MH-53480, MH-53668 and by a grant from the John D and Catherine T MacArthur Foundation.

References

1 Neale MC, Stevenson J. Rater bias in the EASI temperament scales: A twin study. J Personality Soc Psychol 1989; 56: 446–455.
2 Plomin R, Emde R, Braungart JM, Campos J, Corley R, Fulker DW, Kagan J, Reznick S, Robinson J, Zahn-Waxler C, DeFries JC. Genetic change and continuity from 14 to 20 months: The MacArthur Longitudinal Twin Study. Child Devel 1993; 64: 1354–1376.
3 Stevenson J, Fielding J. Ratings of temperament in families of young twins. Br J Devel Psychol 1985; 3: 143–152.
4 Goldsmith HH, Buss KA, Lemery KS. Toddler and childhood temperament: Expanded content, stronger genetic evidence, new evidence for the importance of environment. Devel Psychol 1997; 33: 891–905.
5 Buss AH, Plomin R. Temperament: Early Developing Personality Traits. Erlbaum: Hillsdale, NJ, 1984.
6 Cherny SS, Fulker DW, Corley R, Plomin R, DeFries JC. Continuity and change in infant shyness from 14 to 20 months. Behav Genet 1994; 24: 365–380.
7 Matheny AP Jr. Bayley's Infant Behavior Record: Behavioral components and twin analysis. Child Devel 1980; 51: 1157–1167.
8 Matheny AP Jr. A longitudinal twin study of stability of components from Bayley's Infant Behavior Record. Child Devel 1983; 54: 356–360.
9 Robinson JL, Kagan J, Reznick JS, Corley R. The heritability of inhibited and uninhibited behavior: A twin study. Devel Psychol 1992; 28: 1030–1037.
10 Saudino KJ, Eaton WO. Infant temperament and genetics: An objective twin study of motor activity level. Child Devel 1991; 62: 1167–1174.
11 Saudino KJ, Eaton WO. Continuity and change in objectively assessed temperament: A longitudinal twin study of activity level. Br J Devel Psychol 1995; 13: 81–95.
12 Emde RN, Hewitt JK. The Transition from Infancy to Early Childhood: Genetic and Environmental Influences in the MacArthur Longitudinal Twin Study. Oxford University Press: New York, (in press).

13 Saudino KJ, Cherny SS. Parent ratings of temperament in twins. In: Emde RN, Hewitt JK (eds). The Transition from Infancy to Early Childhood: Genetic and Environmental Influences in the MacArthur Longitudinal Twin Study. Oxford University Press: New York, (in press).
14 Saudino KJ, Cherny SS. Sources of continuity and change in observed temperament. In: Emde RN, Hewitt JK (eds). The Transition from Infancy to Early Childhood: Genetic and Environmental Influences in the MacArthur Longitudinal Twin Study. Oxford University Press: New York, (in press).
15 Rowe DC, Plomin R. Temperament in early childhood. J Personality Assess 1977; 41: 150–156.
16 Bayley N. Manual for the Bayley Scales of Infant Development. The Psychological Corporation: New York, 1969.
17 Hubert NC, Wachs TD, Peters-Martin P, Gandour MJ. The study of early temperament: Measurement and conceptual issues. Child Devel 1982; 53: 571–600.
18 Rothbart MK. Longitudinal observation of infant temperament. Devel Psychol 1986; 22: 356–365.
19 Seifer R, Sameroff AJ, Barrett LC, Krafchuck E. Infant temperament measured by multiple observations and mother reports. Child Devel 1994; 65: 1487–1490.
20 Spinath FM, Angleitner A. Contrast effects in Buss and Plomin's EAS questionnaire: a behavioral–genetic study on early developing personality traits assessed through parental ratings. Personality Individ Diff 1998; 25: 947–963.
21 Carey G. Sibling imitation and contrast effects. Behav Genet 1986; 16: 319–341.
22 Neale MC, Cardon LR. Methodology for Genetic Studies of Twins and Families. Kluwer Academic Publishers: Boston, MA, 1992.
23 Reznick SJ, Robinson J, Corley R. A longitudinal twin study of intelligence in the second year. Monographs of the Society of Research in Child Development Serial No 249, Vol 2, No 1, 1997.
24 Epstein S. Aggregation and beyond: Some basic issues on the prediction of behavior. J Personality 1983; 51: 360–392.
25 Neale M. Mx: Statistical Modeling, 4th edn. Medical College of Virginia, Department of Psychiatry: Richmond, VA, 1997.
26 Plomin R, DeFries JC, McClearn GE, Rutter M. Behavioral Genetics, 3rd edn. WH Freeman: New York, 1997.
27 Eaves LJ. A model for sibling effects in man. Heredity 1976; 36: 205–214.
28 Buss AH, Plomin R, Willerman L. The inheritance of temperaments. J Personality 1973; 41: 513–524.
29 Cyphers LH, Phillips K, Fulker DW, Mrazek DA. Twin temperament during the transition from infancy to early childhood. J Am Acad Child Adolesc Psychiatr 1990; 29: 392–397.
30 Emde RN, Plomin R, Robinson J, Reznick JS, Campos J, Corley R, DeFries JC, Fulker DW, Kagan J, Zahn-Waxler C. Temperament, emotion, and cognition at 14 months: The MacArthur Longitudinal Twin Study. Child Devel 1992; 63: 1437–1435.
31 Torgersen AM. Genetic factors in temperamental individuality A longitudinal study of same-sexed twins from two months to six years of age. J Am Acad Child Psychiatr 1981; 20: 702–711.
32 Bates JE, Bayles K. Objective and subjective components in mothers' perceptions of their children from age 6 months to 3 years. Merrill-Palmer Quarterly 1984; 30: 111–130.
33 Bates JE, Freeland CA, Lounsbury ML. Measure of infant difficultness. Child Devel 1979; 50: 794–803.
34 Daniels D, Plomin R, Greenhalgh J. Correlates of difficult temperament in infancy. Child Devel 1984; 55: 1184–1194.
35 Vaughn BE, Bradley CF, Joffe LS, Seifer F, Barglow P. Maternal characteristics measured prenatally predict ratings of temperamental difficulty on the Cary Infant Temperament Questionnaire. Devel Psychol 1987; 23: 152–161.
36 Mebert CJ. Stability and change in parents' perceptions of infant temperament: Early pregnancy to 13.5 months postpartum. Infant Behav Devel 1989; 12: 237–244.
37 Mebert CJ. Dimensions of subjectivity in parents' ratings of infant temperament. Child Devel 1991; 62: 352–361.
38 Wolk S, Zenah CH, Garcia Coll CT, Carr S. Factors affecting parents' perceptions of temperament in early infancy. Am J Orthopsychiatr 1992; 62: 71–82.
39 Zenah CH, Keener MA, Stewart L, Anders TF. Prenatal perception of infant personality: A preliminary investigation. J Am Acad Child Psychiatr 1985; 24: 204–210.
40 Plomin R, DeFries JC. Origins of individual differences in infancy: The Colorado Adoption Project. Academic Press: Toronto, 1985.
41 Plomin R, DeFries JC, Fulker DW. Nature and Nurture During Infancy and Early Childhood. Harvard University Press: New York, 1988.
42 DeFries JC, Plomin R, Fulker DW. Nature and Nurture during Middle Childhood. Blackwell: Cambridge, MA, 1994.
43 Plomin R, Coon H, Carey G, DeFries JC, Fulker D. Parent–offspring and sibling adoption analyses of parental ratings of temperament in infancy and early childhood. J Personality 1991; 59: 705–732.
44 Schmitz S. Personality and temperament. In: DeFries JC, Plomin R, Fulker DW (eds). Nature and Nurture during Middle Childhood. Blackwell: Cambridge, MA, 1994.
45 Saudino KJ, Gagne JR. Genetic analysis of parent, teacher, and self-reported temperament in early adolescence. Paper presented at the 10th Annual Convention of the American Psychological Society, Washington, DC, May 1998.
46 Schachter FF, Gilutz G, Shore E, Adler M. Sibling deidentification judged by mothers: Cross-validation and developmental studies. Child Devel 1978; 49: 543–546.
47 Schachter FF, Shore E, Feldman-Rotman S, Marquis RE, Campbell S. Sibling deidentification. Devel Psychol 1976; 12: 418–427.
48 Schachter FF, Stone RK. Difficult sibling, easy sibling: Temperament and the within-family environment. Child Devel 1985; 56: 1335–1344.
49 Schachter FF, Stone RK. Comparing and contrasting siblings: Defining the self. J Children Contemp Soc 1987; 19: 55–75.