Acta Genet Med Gemellol 33: 181-189 (1984)

TWIN RESEARCH 4 - Part B: Twin Psychology and Behavior Genetics
Proceedings of the Fourth International Congress on Twin Studies (London 1983)

Twin Similarity in the Developmental Transformations of Infant Temperament as Measured in a Multi-Method, Longitudinal Study

Adam P. Matheny, Jr.
Louisville Twin Study, Department of Pediatrics, University of Louisville School of Medicine, Louisville, Kentucky

Abstract. In the Louisville Twin Study, laboratory observations of twin infants' temperament at 12, 18, and 24 months were linked with parental ratings from temperament questionnaires. Core dimensions of temperament were extracted by factor analysis applied to each set of measures at each age. The laboratory temperament dimension was recurrently represented by emotional tone, social orientation, attentiveness, and reaction to restraint. The questionnaire temperament dimension was recurrently represented by mood, approach/withdrawal and adaptability. The laboratory and questionnaire dimensions were found to be correlated at each age (convergent correlations: 0.38 to 0.52) and to be stable across ages (stability correlations 0.37 to 0.66). The temperament dimensions were used to demonstrate that temperament profiles were more concordant for identical than for fraternal twin pairs. The results demonstrate the genetic influences on (a) the primary dimensions of temperament and (b) the developmental transformations of temperament.

Key words: Infant temperament, Longitudinal study, Direct observations, Parental ratings, Factor validity and stability, Twin concordance

INTRODUCTION

For more than two decades the Louisville Twin Study has examined the genetic influences upon the mental and physical development of twins. More recently, the research program has redirected its efforts to include a longitudinal assessment of infant behavioral characteristics typically described as temperament. That redirection came about because, from our study, intrinsic patterns of behavior were consistently evident as vital features of a twin's individuality. Within each of the twin pairs, contrasts for those patterns could be detected by the parents when the twins were infants, and each twin's pattern became an

identifiable marker for differences in temperament. Thus, it became clear that twins were a convenient means to study the intrinsic origins of temperament.

Our first appraisal of the emergence and stability of rudimentary patterns of temperament was based on parental interviews that elicited the parents' perceptions of behavioral similarities and differences within the twin pair. Data from those interviews indicated that one infant twin, in contrast with the other, could be characterized as more emotionally labile: prone to outbursts of temper, irritable, and demanding of attention. In contrast, the cotwin was seen as less upset during daily routines and more content during self-sustained play with toys. That sharp contrast remained as a stable feature of the intrapair behavioral differences throughout the first three years of life.

During the same developmental interval, the parents made a further differentiation between the twins. Within-pair comparison represented one twin as being more approachful to strangers and smiling more readily, in contrast with the more somber, withdrawing cotwin.

It was evident that data from parental interviews depicted two identifiable trends: first, the early emergence and persistence of a cluster representing temperamental reactivity, and second, a later emergent and largely independent cluster representing social responsiveness. In addition, twin analyses suggested that the core variables comprising the temperamental cluster, and to a lesser extent the social cluster, were influenced genetically [8,15].

Those analyses, however, were based only on intrapair contrasts, a method that does not permit an estimate of the magnitude of contrasts among all sets of twins. Therefore, we turned to methods which would permit a more complete appraisal of the developmental patterns of temperament.

Lytton [4] had previously evaluated the strengths and weaknesses of different methods (eg, direct observations, parental reports) for assessing children's behavior. In view of that evaluation, we did not feel that one method had a clear advantage over any other. Consequently, we undertook a longitudinal appraisal of temperament by concentrating on two sources of data: direct observations obtained within a laboratory setting, and parental reports obtained from temperament questionnaires.

We chose direct observations by trained observers because that course provided more reliable measures. In addition, the observations were made in a structured laboratory setting which provided equivalent conditions within, as well as between, twin pairs. By necessity, however, the periods of observations were short and the variety of situations was constrained.

On the other hand, parental questionnaires have provided a great deal of data representing a wider scope of observations. Parents can report temperament behaviors for events, situations and time spans not readily available during the short looks by trained observers. However, such reports have been criticized on the grounds that they are parental perceptions subject to whatever distortion may occur when parents depict the behavior of their offspring [3].

Our decision to use both methods in order to assess temperament required several steps: (1) developing the structured laboratory observations, (2) selecting a series of appropriate temperament questionnaires, (3) extracting the core dimensions of temperament from each source of data, (4) determining the convergence (if any) between those dimensions at each age, (5) determining the stability and transformation of those dimensions over several ages, and (6) examining the genetic influence on dimensions of temperament as revealed by analyses for twin pairs.

Specific details regarding the steps listed above have been provided in several reports [5,6,7,16]. To date, the most complete sets of data have been obtained for twins at three ages: 12, 18, and 24 months. Analyses of those data have identified the core dimensions from the laboratory observations and questionnaires at each age. In addition, the stability of those dimensions has been established for the same span of ages [7].

With a summary of the previous reports as background, the major purpose of this paper is to present preliminary results from twin analyses applied to the core dimensions of temperament previously identified.

MATERIALS AND METHODS

Subjects

The infants in this study were twins recruited as part of a longitudinal study of twins. At the time of this report, data on twin pairs with identified zygosity were available for 23 pairs (13 MZ, 10 DZ) of 12-month-old twins, 38 pairs (22 MZ, 16 DZ) of 18-month-old twins, and 39 pairs (21 MZ, 18 DZ) of 24-month-old twins. The number of pairs available at each age varied because of missing visits and the ongoing recruiting within the longitudinal study.

The infants were recruited from families in the metropolitan Louisville area. Occupations of head of household, as converted to Duncan's scores for socioeconomic status [10], represented the entire distribution of socioeconomic status, with 27% of the families in the lowest two deciles of the 100-point scale. The remaining families were represented in about equal proportions among the remaining eight deciles.

Recruited twins made quarterly visits to the research center during the first year, and visits every six months thereafter until age 3 years. It was not until 12 months and thereafter that the full range of laboratory measures could be employed. Consequently, the present study was based on data from the largest samples available from visits incorporating the widest range of measures. Those visits were for the ages 12, 18, and 24 months.

Temperament Questionnaire

Parental reports on infant temperament were obtained from the *Toddler Temperament Scale (TTS)* which consists of 97 items rated on 6-point scales [2]. Sets of items are combined to yield nine scores representing the nine categories of temperament proposed by reports from the New York Longitudinal Study [11,13]. Those nine categories are given as the following: (1) activity, (2) rhythmicity, (3) approach/withdrawal, (4) adaptability, (5) intensity of reaction, (6) mood, (7) attention/persistence, (8) distractibility, and (9) threshold of responsiveness.

The *TTS* was chosen for parents to complete because it samples a wide variety of commonplace events occurring to infants between one and three years of age. The median for internal consistency reliability is reported to be 0.70. Median test-retest reliability is 0.81 for the nine scales given twice with an interval of one month.

Laboratory Observations and Rating Scales

During visits to the laboratory, the infants were engaged in a standard set of specific activities organized in a prearranged sequence in a playroom. The activities, called vignettes, generally took place without the presence of the infants' mothers and with (1) each twin infant engaged alone with a staff member, or (2) both twin infants together and engaged individually with two staff members. All vignettes were scheduled in a sequence exactly duplicated for all infants. Videotapes of the vignettes were made according to a format carefully organized to yield one hour of videotape representing a morning visit at the laboratory. The organization of the visits, schedule and description of the vignettes, and format for videotaping have been described fully elsewhere [6]. In all, there were 15 activities provided, representing 30 minutes of videotaping time for each twin.

The rating scales developed for use in the laboratory were derived from the behavioral categories of the *Infant Behavior Record* [1] which had been extensively used and refined in the longitudinal study [5]. The laboratory observations were rated on 9-point rating scales representing the more descriptively prominent scales from the *Infant Behavior Record:* Emotional tone, Attention span, Activity level, Social orientation to staff member, and Vocalizing.

Ratings from a videotape were made by having raters – who had not interacted with the infants – independently view the videotape and then make the ratings for each successive 2-minute period. As a consequence, for each infant there were 15 ratings made on each of the five rating scales over the course of observations.

In addition to the ratings made from the videotapes, 9-point ratings were recorded *in situ* for the infants' emotional tone, cooperation, and resistance to restraint when physical measurements were made. During the last part of a visit, infants were taken to a small adjacent room where they were undressed, restrained for measures of head circumference and weight, and further restrained while being stretched out for measures of reclining length.

Preliminary analyses indicated that the ratings made during physical measures were highly intercorrelated; therefore, a single score, representing the sum of the ratings was derived. That derived score was called Reaction to restraint.

Scoring and Factor Scores

Questionnaire. The scores for the nine categories of temperament were subjected to principal components factor analysis at each age. Both unrotated and rotated solutions were examined, and the unrotated solutions were retained for interpretation.

The first factor was particularly interesting, in that it was recurrently defined by the same temperament categories at every age. The temperament categories, mood, adaptability, and approach/withdrawal primarily anchored the first factor, thereby contributing to a salient temperament dimension as seen by the parents. At one extreme, infants were seen as positive in mood, adaptable to changes and approachful to new people and events; at the other, infants were seen as irritable and negative, slow to adapt, and avoidant of new persons or situations.

In general, the first factor appeared to incorporate the distinction between easy vs difficult temperament made by Thomas et al [12]. At every age, the data supported the impression that there is a basic temperament cluster reflecting a dimension of tractability vs intractability. Therefore, the first factors – labeled Questionnaire-Tractability – from the questionnaires at 12, 18 and 24 months became part of the core dimensions of temperament later used for the twin analyses.

Laboratory. After each visit was completed, raters worked individually from the videotape and made the ratings for each scale for each successive 2-minute period. This procedure generated 15 ratings for each of the five rating scales used in the playroom. In order to condense the large matrix of ratings, a single score was created for each of the five rating scales. That score represented the average value of the ratings on a scale over all of the rating periods. Another score was created in order to represent the extent of change or variability in activity over the 15 periods.

As a consequence of the condensation procedures, 6 scores became the bulk of the primary data for each infant's temperament as observed in the playroom. To these scores was added the score for reaction to restraint obtained during physical measures. Thus, the laboratory observations at 12, 18, and 24 months consisted of the following measures:

1. Emotional Tone
2. Activity
3. Activity (Variability)
4. Attentiveness
5. Social Orientation: Staff
6. Vocalizing
7. Reaction to restraint

Correlations among the laboratory measures at each age indicated that a core cluster of measures recurrently occurred. That cluster consisted of emotional tone, attentiveness, social orientation to staff, and reaction to restraint. As a further step in defining the cluster, a principal-components factor analysis was performed at each age and two factors with eigenvalues >1.00 were found. The rotated and unrotated solutions were quite similar: however, the unrotated solutions more clearly emphasized the core cluster identified in the correlation matrices.

The first factor, structurally similar at every age, was composed of high positive loadings from emotional tone, attentiveness, social orientation to staff and reaction to restraint, as well as a negative loading from activity, variability. Infants with high scores on this factor could be described as positive in affect, approachful to staff, capable of sustained attention during the task sequences, acceptant of the restraints during physical measures and less labile in activity level from one rating period to the next. Infants with low scores were negatively resistant, unresponsive to or wary of the staff, inattentive, upset during physical measures, and erratic in activity. These two extremes, as determined by

the laboratory procedures, represented a dimension of tractability in the laboratory. Accordingly, the first laboratory factor at each age was labeled Lab-Tractability and it became the core dimension of temperament for the direct observations of infant temperament.

RESULTS

The products of the foregoing procedures for data reduction permitted each infant to be characterized by a location on two dimensions (factors) at each age. One location represented the infant's temperament as shown by a first-factor score from the laboratory observation: Lab-Tractability. The other location represented the infant's temperament as shown by the first factor of the TTS: Questionnaire-Tractability.

Convergent Validity and Stability

The complete analyses demonstrating the convergence between Lab-Tractability and Questionnaire-Tractability have been reported elsewhere [7,16]. Those analyses showed that there was a significant overlap between the core dimensions of temperament as extracted from direct observations and parental reports. In spite of the differences in the objectivity of observers, methods of measurement, and periods of observation, a consistency of temperament was reliably detected over both settings. Infants observed in the laboratory to be more positive, attentive, socially oriented and more placid during physical measures were described by their mothers as more positive, adaptable, and approachful – a convergent relation that was confirmed at all three ages. The convergent correlations depicting these relations are provided in Table 1.

Table 1 also shows the degree of stability in temperament over successive ages. Inspection of the correlations indicates that both the laboratory and questionnaire first-factors are moderately stable from 12 to 18 months, and increase in stability from 18 to 24 months. Apparently, there was considerable reordering of individual differences for the two measures of temperament during the first six-month interval and a consolidation of the ordering of individual differences during the second six-month interval. In effect, for the individual infants, there was some degree of transformation of temperament taking place from 12 to 24 months. However, the collective relations, across settings and over ages, indicated that an underlying dimension of temperament was retained. The repetition of the key variables, the replications of the factor structures, and the convergence among data sets are consistent with a view that there is a distinctive pattern of temperament for each infant in the second year of life.

Twin Analyses

With the results from individual twin infants as a background, the next question concerns the degree of genetic influence on the core nucleus of temperament. In this instance, the temperament scores for the twin infants, reconstituted into MZ and DZ twin pairs, would become the means for demonstrating that genetic influence.

For the twin analyses, the core nucleus of temperament was represented by a single standardized score for each twin at each age. That general temperament score was obtained by adding the first-factor scores from the laboratory and questionnaire and standardizing the combined scores.

TABLE 1 - Primary Temperament Factors: 12, 18, and 24 Months

Source of observations	Primary factors					
	Behaviors (loading)		Behaviors (loading)		Behaviors (loading)	
Laboratory	Emotional tone (0.86) Attentiveness (0.84) Social orientation: staff (0.65) Reaction to restraint Emotional tone (0.62) Cooperation (0.59)	r = 0.37	Emotional tone (0.92) Attentiveness (0.87) Reaction to restraint (0.71) Social orientation: staff (0.64) Activity: variability (−0.55)	r = 0.66	Emotional tone (0.94) Attentiveness (0.92) Social orientation: staff (0.84) Activity: variability (−0.66) Reaction to restraint (0.55)	
	r = 0.52		r = 0.38		r = 0.52	
Toddler Temperament Scale	Adaptability (0.84) Attention/persistence (0.80) Approach/withdrawal (0.74) Mood (0.73) Distractibility (−0.65)	r = 0.44	Mood (0.86) Adaptability (0.75) Intensity (−0.69) Distractibility (−0.53) Approach/withdrawal (0.52)	r = 0.65	Mood (0.82) Adaptability (0.76) Approach/withdrawal (0.64) Activity (−0.55) Intensity (−0.54)	
	12 Months		18 Months		24 Months	
	Age					

Behaviors with loadings ≥ 0.50 are shown. All correlations are significant, $P \leq 0.05$ (N's vary between 37 and 48).
The tabular material is adapted from Matheny et al [7].

TABLE 2 - Twin Similarity for General Temperament at 12, 18, and 24 Months

Age	Within-pair correlations		
	MZ	DZ	No. of pairs
12 months	0.71	0.45	13-10
18 months	0.88*	0.17	22-16
24 months	0.84	0.76	21-18

* $P<0.05$ for $R_{MZ}>R_{DZ}$

TABLE 3 - Twin Similarity for Temperament: Developmental Trend

Age span	Profile R		Age-to-age correlation: all twins
	MZ	DZ	
12-18 months	0.77*	0.12	0.50
18-24 months	0.66	0.65	0.63
12-18-24 months	0.52	0.07	0.57

Number of pairs for MZ/DZ: 12:18 months,11/8; 18-24 months, 13/8; 12-18-24 months, 10/6.
* $P<0.05$ for $R_{MZ}>R_{DZ}$.

At each age, the concordance within the MZ and DZ twin pairs for the general temperament score was determined by computing intraclass (within-pair) correlations for the scores at each age. The results are presented separately for the MZ and DZ pairs in Table 2.

The results show that the MZ correlations were higher than the DZ correlations at every age; however, the differences were only significant at 18 months.

Apparently, there is some degree of genetic influence on the core nucleus of temperament as measured during the second year. Yet, the previous results indicated that from age to age, there was some degree of reordering of the individual differences in temperament during that same period. Therefore, the next analysis turned to the question of MZ and DZ concordance for change in general temperament from one age to the next.

By relying again on the single scores representing general temperament, the longitudinal analyses were obtained through an analysis of variance specifically adapted for twin data involving congruence in longitudinal profiles [14]. Here the analyses were based on each twin's profile of temperament scores across ages, and the degree to which the twins within pairs have congruent profiles.

The profile correlations, shown in Table 3, are particularly interesting. They indicate that the MZ twin pairs have profile correlations higher than the DZ twin pairs. If one recalls that the temperament scores undergo some degree of change, as represented by the age-to-age correlations in Table 3, then the results are clear; transformations in temperament can be attributed to some degree of genetic influence.

DISCUSSION

It is evident that multi-method assessments of temperament among infants between 12 and 24 months can lead to the isolation of a general dimension of temperament. That dimension underlies several categories of temperament rated by parents and similar aspects of temperament rated in a laboratory setting. More specifically, that dimension repeatedly bridges both sets of measures for aspects of an infant's positive or negative mood, approach or withdrawal from people and novel events, and adaptability or the lack of it to new or altered situations.

The replication of the temperament structures for either set of measures at three ages and the repetition in convergence between those two sets are striking findings. They are further strenghtened by the evidence that the temperament structures are moderately stable during a developmental period in which the rapid deployment of locomotor skills and language may alter the form and substance of temperament. The aggregate results point to the distinctive coherency of general temperament as crystallized from two sets of data.

Theories of temperament have typically emphasized the potential genetic influence on the emergence and maintenance of aspects of temperament. Twin studies have addressed some of the issues regarding the type and specificity of temperament(s) influenced genetically; however the findings have not yet resolved how pervasive genetic influences might be. For example, genetic influences have been found for a particular set of temperament characteristics at one age; at another age genetic influences may not be found for that set but another. To complicate matters even further, parental reports of twins' temperament typically provide strong and pervasive evidence of genetic influences on temperament while direct observations provide evidence that genetic influences are weaker and more sharply differentiated. For example, parental reports suggest that sociability among children is markedly influenced genetically, but direct observations suggest that the genetic influence is demonstrable for children's behavior toward strangers and not familiars [9]. In conjunction with additional evidence that the identity of temperament behaviors may not be stably predicted from one age to the next, the scope of genetic influences on temperament remains unestablished.

In some respects, the constraints for research on the genetic influences upon temperament are those for research on personality in general. The generalities of temperament across settings, intervals of time, and sources of observation may always be quite modest because of situational variables which have some bearing on the types as well as the range of individual differences observed. For example, infants placid and sociable in the home situation may be seen as more distraught and withdrawn in the laboratory setting. Also, the same discrepancies among the characteristics of temperament occur when one compares observations at one age with observations at the next age. Recognition of these situational contingencies argues for an empirical approach that distills the core dimensions that may transcend settings or time.

Despite the development of empirically refined approaches to assessing temperament, we should continue to expect transformations in temperament. A number of factors, not the least of which is the individual variations in rates of maturation, may confound efforts to demonstrate consistency of patterns of temperament for all occasions. The twin analyses, however, suggest that there is a genetic influence on the changes themselves. Although the age-to-age profile correlations for twin pairs should be viewed with caution

because of the small samples, the correlations depict transformations of temperament as being subject to an underlying regularity. Thus, the changes that occur seem to be mediated through the infant's innate tendencies which both permit and restrict individual variations across occasions.

These preliminary findings require support from data based on larger samples and a wider span of ages; however, they illustrate the potential explanatory advantages from examining behavior-genetic relations within a developmental context. It is anticipated that the developmental organization of temperament, as well as the nominal discontinuities of temperament, may have intrinsic determinants as influential as those found for other developmental characteristics.

Acknowledgments. Supported by research grant 90-C-922 from the Office of Child Development, and a grant from the John D. and Catherine T. MacArthur Foundation (Ronald S. Wilson, principal investigator). I am indebted to the many coworkers who have made contributions to the research program, including R. Arbegust, P. Gefert, M. Hinkle, J. Lechleiter, S. Nuss, and B. Moss.

REFERENCES

1. Bayley N (1969): Bayley Scales of Infant Development. New York: Psychological Corporation.
2. Fullard W, McDevitt SC, Carey WB (1984): Assessing temperament in one to three year old children. J Pediat Psychol, in press.
3. Hubert NC, Wachs TD, Peters-Martin P, Gandour MJ (1982): The study of early temperament: measurement and conceptual issues. Child Dev 53:571-600.
4. Lytton H (1974): Comparative yield of three data sources in the study of parent-child interaction. Merrill-Palmer Quart 20:53-64.
5. Matheny AP (1981): Bayley's infant behavior record: behavioral components and twin analyses. Child Dev 51:1157-1167.
6. Matheny AP, Wilson RS (1981): Developmental tasks and rating scales for the laboratory assessment of infant temperament. JSAS Catalog Selected Documents Psychol. 11:81-82 (MS No 2367).
7. Matheny AP, Wilson RS, Nuss SN (1984): Toddler temperament: stability across settings and over ages. Child Dev, in press.
8. Matheny AP, Wilson RS, Dolan AM, Krantz JZ (1981). Behavioral contrasts in twinships: stability and patterns of differences in childhood. Child Dev 52:579-588.
9. Plomin R, Rowe DC (1979): Genetic and environmental etiology of social behavior in infancy. Dev Psychol 15:62-72.
10. Reiss AJ (1961): Occupations and Social Status. New York: Free Press.
11. Thomas A, Chess S (1977): Temperament and Development. New York: Brunner/Mazel.
12. Thomas A, Chess S, Birch HG (1968): Temperament and Behavior Disorders in Children. New York: New York Univ. Press.
13. Thomas A, Chess S, Birch HG, Hertzig ME, Korn S (1963): Behavioral Individuality in Early Childhood. New York: New York Univ. Press.
14. Wilson RS (1979): Analysis of longitudinal twin data. Acta Genet Med Gemellol 28:93-105.
15. Wilson RS, Brown AM, Matheny AP (1971): Emergence and persistence of behavioral differences in twins. Child Dev 42:1381-1398.
16. Wilson RS, Matheny AP (1983) Assessment of temperament in infant twins. Dev Psychol 19: 172-183.

Correspondence: Dr. Adam P. Matheny, Jr., Louisville Twin Study, Department of Pediatrics, University of Louisville School of Medicine, Louisville, KY 40292, USA.