Genetic Influences on Activity Level:

An Analysis of Continuity and Change from Infancy to Early Childhood

by

Kimberly Jane Saudino

A thesis
presented to the University of Manitoba
in fulfillment of the
thesis requirement for the degree of
DOCTOR OF PHILOSOPHY
in
Psychology

Winnipeg, Manitoba

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services Branch

Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-315-77817-2

Canada

GENETIC INFLUENCES ON ACTIVITY LEVEL: AN ANALYSIS OF CONTINUITY AND CHANGE FROM INFANCY TO EARLY CHILDHOOD

BY

KIMBERLY JANE SAUDINO

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

I hereby declare that I am the sole author of this thesis.

I authorize the University of Manitoba to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Kimberly Jane Saudino

I further authorize the University of Manitoba to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Kimberly Jane Saudino

ABSTRACT

Because genetic influences may not be stable across the lifespan, the finding of genetic influences on activity level (AL) at one age tells little about such influences at another. Previous research exploring the importance of genetic contributions to individual differences in motor activity across age have been clouded by subjective errors that may arise when assessments of AL rely on human judgment. The present study is a longitudinal extension of an initial study demonstrating genetic influences on mechanically-assessed AL in 7-month-old twins (Saudino & Eaton, 1991). The motor activity of 36-month-old twins was re-evaluated with motion recorders and parent ratings over a two-day period. Mechanically-assessed AL continued to show a genetic influence during early childhood ($R_{MZ} = .76$, $R_{DZ} = .38$). Moreover, MZ co-twins displayed greater concordance for *change* in AL than did DZ co-twins ($R_{MZ} = .84$, $R_{DZ} = .48$), thus demonstrating a genetic influence on developmental change.

ACKNOWLEDGEMENTS

There are many people to whom I would like to express my sincere gratitude for their assistance and support throughout this project. First and foremost, thank you to my advisor, Dr. Warren Eaton who has been generous with his time, knowledge, and good humor. Warren's guidance and encouragement has enabled me to achieve things I never thought possible. Thank you also to my committee members Dr. Marion Aftanas, Dr. Gordon Barnes, and Dr. Ab Chudley. Their insightful comments and suggestions added greatly to the quality of this thesis. The contribution of participating families is gratefully acknowledged. Thank you to Marni Brownell for her reliability and reality checks. I would also like to thank my husband, Jeff Zapfe for his continued support and encouragement throughout my graduate career. I only hope I can live up to his example. Finally, thank you to my parents for providing me with great genes and a good environment.

CONTENTS

ABSTRACT iv

ACKNOWLEDGEMENTS v

INTRODUCTION 1
- Developmental Behavioral Genetics 1
- Genetic Influences on AL 3
  - The Classical Twin Design 4
  - Initial Twin Study 7
- Genetic Influences on AL: A Developmental Perspective 13
  - Fundamental Questions 13
    - Differential Heritability Across Age 13
    - Genetic Contributions to Developmental Change 15
      - Individual differences and development 16
      - Assessing the role of genes in developmental change 16
      - Assessing the role of genes in developmental continuity 17
    - Longitudinal Versus Cross-sectional Designs 18
  - Empirical Findings 19
    - Cross-sectional Studies 19
    - Longitudinal Studies 23
    - Conclusions 33
  - The Present Study 39
    - Hypotheses 41

METHOD 44
- Participants 44
  - Recruitment 44
  - Sample 44
    - Analysis for Possible Selective Attrition 45
- Procedure 46
  - Overview 46
  - Instruments 47
    - Actometer Measure 47
    - Motor Development Measure 49
    - Parent Activity Questionnaire 50
    - Physical Measures 51

Zygosity 51
Reliability of Physical Similarity Judgments 53
Validity of Zygosity Diagnoses 53
Twin Analyses 55
Adjusted Activity Scores 55
Preliminary Analyses 56
Intraclass Correlations 56
Genetic Analysis of Change and Continuity 58
Twin Concordance for Change 58
Cross-twin Intraclass Correlations 59

RESULTS 60
Descriptive Statistics 60
Phenotypic Stability 63
Genetic Influences at 7 and 36 Months 64
Actometer-Assessed AL 65
Composite Actometer Score 66
Adjusting for Age 67
Adjusting for Motor Maturity 67
Parent-Rated AL 68
Motor Development 69
Anthropometric Measures 69
Differential Heritability 70
Genetic Influences on Change and Continuity 70
Analysis of Change 70
Analysis of Continuity 72

DISCUSSION 76

REFERENCES 89

APPENDICES

APPENDIX A: Subject Recruitment Letter 102

APPENDIX B: Telephone Protocol 104

APPENDIX C: Consent to Participate 105

APPENDIX D: Demographic Questionnaire 106

APPENDIX E: Actometer Instructions 108

APPENDIX F: Actometer Record Sheet (Parent) 109

APPENDIX G: Instrument Measurement Properties 110

APPENDIX H: Summary of Derived Variables 119

APPENDIX I: The TBAQ Activity Subscale 120

APPENDIX J: The Diagnosis of Zygosity 123

APPENDIX K: Zygosity Questionnaires 126

APPENDIX L: Zygosity Decision Rules 129

LIST OF TABLES

Table 1. Twin Intraclass Correlations for Activity Factor Adapted From Matheny (1980) 26

Table 2. Parent-researcher Agreement for Judgments of Physical Similarity 54

Table 3. Means and Standard Deviations for Core Variables at 7 and 36 Months of Age 61

Table 4. Significant *F* Statistics for Age and Sex Effects by Twin 62

Table 5. Phenotypic Stability Correlations from 7 to 36 Months of Age . . . 64

Table 6. Intraclass Correlations, Standard Errors, and Statistical Tests of Genetic Variance at 7 and 36 Months of Age 65

Table 7. Intraclass Correlations, Standard Errors, and Statistical Tests of Genetic Variance for Change and Slope Scores from 7 and 36 Months of Age 72

Table 8. Cross-twin Correlations and Standard Errors for Genetic Continuity From 7 to 36 Months of Age 74

Genetic Influences on Activity Level:

An Analysis of Continuity and Change from Infancy to Early Childhood

Despite the diversity of child temperament theories, most contemporary approaches acknowledge the importance of genetic influences on individual differences in behavior. As a consequence, there has been a profusion of behavioral genetic studies examining the etiology of temperament variability (see Goldsmith, 1983, for a review). Such studies attempt to determine the degree to which genetic and environmental influences contribute to observed individual differences in those behaviors thought to reflect temperamental dimensions.

While the value of this research strategy is obvious, as McCall (1986) notes, the genetics of temperament has rarely been studied from a developmental standpoint. That is, few studies explore questions regarding the extent to which genetic and environmental influences on temperament *change* during the course of development or, further, the role that genetic factors may play in *promoting* behavioral development. This is the essence of a new field of study -- *developmental* behavior genetics.

Developmental Behavioral Genetics

The factors that govern individual differences in temperament *within* an age may differ *across* age (McCall, 1986). That is, there may be changes in the relative contribution of genetic and environmental influences throughout the lifespan. Previously, genetic influences were considered immutable and viewed as constants in the developmental process. This led to the mistaken notion that only longitudinally stable characteristics were genetically determined, whereas unstable, malleable characteristics were the result of environmental factors (Plomin, 1983). Consistent with this notion is the "hypothesis of decreasing magnitude of heritability" (Plomin, et al., 1992). This hypothesis suggests that as a child matures, encounters more diverse environments, and gains greater control in interacting with them, the impact of environmental factors may become more salient for certain traits. The resulting increase in environmental variance in tandem with stable genetic variance would result in decreased heritability, the proportion of total variance attributable to genetic factors.

Contrary to this view, genes are, in fact, dynamic in nature, with certain genes switching on and off, or having their maximum influence at different developmental periods. Even when a trait is phenotypically stable over time and age, genes may exert a significant influence at one age but not at another. Similarly, for any trait that displays genetic influences across age, the genes that operate at one age, may differ from those that operate at another.

The focus of developmental behavioral genetics is not, however, simply on how much a trait or ability is influenced by genetic and environmental factors at different periods in the lifespan. The dynamic nature of gene action means that genetic factors can induce developmental change. Thus, a fundamental proposition of developmental behavior genetics is that genes are considered as potential sources of both change and continuity in development (Plomin, DeFries, & Fulker, 1988).

Both stability and developmental change are considered in the present research, which is a longitudinal extension of an initial study exploring the role of genetic influences on objectively assessed activity level in infant twins (Saudino & Eaton, 1991). This study examines, from a truly developmental perspective, the importance of genetic contributions to individual differences in motor activity. Activity level (AL) is a core dimension of nearly every temperament theory (Goldsmith et al., 1987; Hubert, Wachs, Peters-Martin, & Gandour, 1982). Defined as an individual's customary level of energy expenditure through gross motor movement, it is frequently studied and, perhaps, the best validated dimension of temperament. It has been demonstrated to be a significant individual difference variable in both infants and preschoolers (Eaton, 1983; Eaton & Dureski, 1986) and appears to show continuity in expression across time and situations (Hubert et al., 1982). Moreover, activity level is one temperament dimension for which there exists considerable empirical evidence of a genetic influence (e.g, Buss, Plomin, & Willerman, 1973; Saudino & Eaton, 1991; Stevenson & Fielding, 1985; Torgersen, 1981; Willerman, 1973).

Genetic Influences on AL

An integral concept of behavioral genetics is that the behavior of individuals in the population varies due to both genetic and environmental influences. To explore the degree to which genetic and environmental variance covary with behavioral variability, the behavioral geneticist studies individuals who vary systematically in their genetic and/or environmental similarity. These individuals are typically included in twin studies, adoption studies and family studies.

The strongest evidence for genetic influences on activity level comes from research employing the twin study design. By comparison, there is a relative lack of evidence from family and adoption studies. Presumably, this results from the problems that arise when making cross-age comparisons on temperament traits. For example, when evaluating the degree of similarity between a parent and child, the expression of, or salience of, AL might differ at each age. Similarly, across the two age groups there may be a considerable degree of disparity between the measures employed to assess AL, and this would cloud any comparison of the two measurements. Perhaps more important is the possibility of differential heritability across age. For parent-offspring designs to demonstrate significant heritability, the trait under study must be heritable in *both* childhood and adulthood. Such generational differences cannot confound the analysis of twin studies because co-twins are matched precisely for chronological age. Given this consideration, twin studies will probably be the most appropriate method for detecting genetic influences in infancy, a period of rapid developmental change (Goldsmith, 1983).

The Classical Twin Design

The twin study design involves contrasting identical or monozygotic (MZ) twins, with same-sex fraternal, or dizygotic (DZ), twins. Monozygotic co-twins possess identical genotypes, having 100% of their genes in common. Therefore, any differences between MZ co-twins are assumed to reflect both environmental influences that are not shared by the twins (non-shared environment) and measurement error. Dizygotic twins are half as genetically similar as MZ twins, and like all first degree relatives, share, on average, 50% of their segregating genes. Thus, DZ co-twins differ from each other for genetic *and* environmental reasons.

In comparing MZ and DZ twins, it is postulated that should MZ co-twins show more behavioral similarity, it must be a result of genetic influences due to the two-fold greater number of shared genes. Thus, according to genetic theory, if a trait is genetically influenced, MZ co-twins should be approximately twice as similar as DZ co-twins. If, however, MZ co-twins are no more alike than DZ co-twins, the trait under study is considered to show no genetic influence.

Typically, intraclass correlations serve as indices of co-twin similarity and, unlike the more familiar Pearson correlations, represent an estimate of shared variance. Genetic influences are indicated when MZ intraclass correlations are significantly larger than DZ intraclass correlations. An estimate of heritability, the proportion of observed behavioral variance explained by genetic variance, can be derived by doubling the difference between the intraclass correlations for the two twin types (Falconer, 1981).

In order to draw inferences about genetic influence from the classical twin design, three critical assumptions are required:

1. **Environmental influences are equally similar for both twin types.** If this assumption is not met, observed, or phenotypic, differences between MZ and DZ twins would reflect environmental influences in addition to genetic influences (Plomin & DeFries, 1985). Studies examining the equal environments assumption provide evidence that it is a tenable proposition (e.g., Scarr, 1966; Scarr & Carter-Saltzman, 1979; Torgersen & Kringlen, 1978). Although similarity of physical appearance may create some treatment inequalities amongst twin types, it does not appear to significantly bias twin studies (i.e., by inflating heritabilities) that examine personality, cognitive or perceptual abilities (e.g., Plomin, Willerman, & Loehlin, 1976; Matheny, Wilson, & Dolan, 1976). Moreover, some environmental differences between MZ and DZ twins reflect parental response to genetic differences (Lytton, 1977; Scarr & Carter-Saltzman, 1979).

2. **No assortative mating.** Under this assumption, mating in the population is random, and there is no tendency for like to mate with like. Should assortative mating occur, parents would be more genetically similar to each other than would transpire by chance. This would increase the genetic similarity of DZ co-twins. Therefore, assortative mating operates to inflate the DZ intraclass correlations, and thereby reduces the differences between MZ and DZ correlations and results in an underestimate of genetic influence.

3. **Genetic variance is additive.** Most traits are multifactorial; that is, they are influenced by multiple genes and the environment. When genetic variance is

additive, the effect of each gene form, or allele, equally contributes to the phenotypic expression of the trait. Thus, additive genetic variance characterizes genetic influences that "breed true" (Plomin, DeFries, & McClearn, 1990). *Non*additive genetic variance does not breed true, that is, there is not a linear relationship between the amount of genes in common and phenotypic similarity. For example, if there is dominance among alleles at a single locus, or if a trait is influenced by epistasis -- the complex interaction of all alleles across loci (Lykken, 1982), the phenotypic expression does not represent the sum of the average effects alleles. Although MZ twins share all nonadditive genetic effects, DZ twins share only a quarter of genetic variance due to dominance and even less variance due to epistasis (Plomin et al., 1988). Thus, if nonadditive genetic variance is important for a characteristic, the similarity of DZ twins will be less that half that of the MZ twins. Consequently, nonadditive variance will lead to inflated estimates of the genetic influence on that characteristic.

The extent to which these last two assumptions are violated would compromise the conclusions drawn from any twin study. However, because the influences of assortative mating and nonadditive genetic variance work in opposite directions, the twin study design can provide a reasonable first approximation of genetic variance (Plomin & DeFries, 1985).

Initial Twin Study

The initial phase of this research (Saudino & Eaton, 1991) was prompted by the ambiguous empirical support that claims for a constitutional or biological basis of temperament had received. Almost without exception, prior studies exploring genetic influences on temperament had employed subjective observer ratings of behaviors thought to reflect temperamental dimensions in infants and children. With such rating techniques, there had emerged evidence for moderate genetic influences on individual behavioral differences, but the data were equivocal as illustrated below.

Although MZ co-twins were found to be consistently more similar on rated temperament dimensions than DZ co-twins, the degree and pattern of resemblance covaried with the specificity of the rating scale. When ratings required global judgments about a child's behavior (e.g., EASI temperament scales, Buss & Plomin, 1975, 1984; Colorado Child Temperament Inventory [CCTI], Rowe & Plomin, 1977; and scales based on the New York Longitudinal Study [NYLS] protocols, Chess & Thomas, 1977), MZ within-pair resemblance was generally moderate, yet the DZ within-pair resemblance was much lower than would be predicted by genetic theory (Buss et al., 1973; Emde et al., in press; Neale & Stevenson, 1989; Plomin & Rowe, 1977; Stevenson & Fielding, 1985; Torgersen & Kringlen, 1978). Moreover, in some instances, the DZ intraclass correlation was *negative*, an implausible outcome given the conclusion of moderate heritability. Activity level was the temperament dimension for which this unusual pattern of results was particularly evident.

In contrast, the use of specific behavior rating items (e.g., Infant Behavior Questionnaire [IBQ], Werry Activity Scale) typically yielded higher estimates of co-twin similarity for both twin types. Further, these latter results fit the genetic model more appropriately. That is, DZ co-twin similarity was approximately one-half that of MZ similarity (e.g., Cohen, Dibble, & Grawe, 1977; Goldsmith & Campos, 1986; Willerman, 1973).

The presence of nonadditive genetic variance has been presented as a possible explanation for DZ correlations that are low relative to MZ correlations (Emde et al., in press; Plomin, Chipuer, & Loehlin, 1990). Indeed, evidence for the influence of nonadditive genetic variance on activity level has emerged from model fitting analyses of the EAS temperaments in the last half of the lifespan (Plomin, Pedersen, McClearn, Nesselroade, & Bergeman, 1988), but has not been clearly demonstrated in twin studies of temperament in infancy and early childhood. Nonetheless, Plomin, Coon, Carey, DeFries, and Fulker (1991) suggest that the hypothesis of nonadditive genetic variance might be invoked to explain the finding that, in direct contrast to twin studies, sibling adoption analyses of parental ratings of temperament in the Colorado Adoption Project (CAP) showed no genetic influence.

While the possibility of nonadditive genetic variance must be seriously considered, it cannot adequately explain the unusual pattern of negative correlations, nor is it likely that nonadditive genetic variance alone could result in near zero correlations (Plomin et al., 1991). Moreover, the problem of low or negative DZ correlations typically occurs when activity level is assessed via global rating measures. If nonadditive genetic variance is present, it should be evident across all methods of assessment.

The difference in outcomes when global versus specific measures are employed suggests that subjective assessments are prone to perceptual rater biases that may accentuate differences between DZ co-twins (contrast effects), or, alternatively, accentuate similarities between MZ co-twins (assimilation effects). This conclusion is consistent with Neale and Stevenson's (1989) recent finding of significant rater bias in the EASI temperament scales, a global rating scale frequently employed in behavior genetic studies of child temperament. Such biases appear to become increasingly conspicuous as the rating task becomes more general. Thus, rating scales likely reflect observer expectations as well as actual child behavior and, consequently, provide an inadequate evidential base for the constitutional emphasis of temperament theory.

For most dimensions of temperament there are few practical alternatives to observer ratings. Activity level is, however, an exception to this dilemma. One can measure AL mechanically and, thus, can circumvent the biases associated with subjective measurement. A finding of genetic influences with mechanical measures would support the assumption of a constitutional basis to temperament. In addition, the use of objective instruments would address the issue of rater bias versus nonadditive genetic variance as explanations for low DZ intraclass correlations. If nonadditive genetic variance is present, one would expect to the pattern of low DZ correlations to persist when AL is measured mechanically.

Previous twin studies that had included objective measures of AL had been limited by poor measure reliability and validity and provided, at best, weak evidence of a genetic influence (e.g., Lytton, Martin, & Eaves, 1977; Scarr, 1966). Nonetheless, the use of mechanical devices to record AL showed some promise in Plomin and Foch's (1980) study of objectively assessed personality in children. Of the variety of objective behavioral measures employed, only a week-long pedometer measure of AL demonstrated sufficient test-retest reliability. Furthermore, this measure yielded evidence of significant, but slight, genetic influences on AL. However, the very high concordances of *both* MZ and DZ co-twins suggested that environmental factors were substantial (i.e., $R_{MZ} = .99$, $R_{DZ} = .94$). Two explanations for this unusual pattern of resemblance can be entertained. First, the twins (mean age 7 years) presumably played together, and the measured activity of each twin was not independent of the other. Second, because in this study, the pedometer was worn at the waist and recorded only the up and down movements of the trunk, the measure may have been sensitive to between-pair differences and yet too crude to detect more subtle within-pair differences in overall activity, particularly of the limbs.

With the need for a stronger test of the genetic model in mind, we objectively assessed the AL of infant twins (mean age 30.7 weeks) with actometers, mechanical motion recorders shown to be valid and highly reliable for measuring infant AL within the home environment (Eaton & Dureski, 1986; Eaton, McKeen, & Lam, 1988). The intraclass correlations for the 48-hour mechanical measure of motor activity provided evidence of a genetic influence on AL ($R_{MZ} = .76$, $R_{DZ} = .56$, $p < .10$). In addition, there was no indication of contrast or assimilation effects, and, twin concordance conformed more closely to the classic twin model of MZ co-twins being approximately twice as similar as DZ co-twins. However, parent ratings of AL on the IBQ Activity subscale suggested the presence of contrast effects. MZ twins were rated as highly similar, whereas the similarity of DZ twins was not significantly different from zero (R_{MZ} = .82, R_{DZ} = .21). This outcome coupled with the fact that the MZ correlation for parent-rated AL was of a magnitude close to that for actometer-assessed AL, leads us to believe that if there is a bias operating, it is a contrast bias. Contrast effects result from the rater's tendency to contrast one twin with the other, thereby magnifying behavioral differences in the process. However, in our study, the DZ correlation did not *significantly* differ from half the MZ correlation, as would be necessary to clearly show a contrast bias.

Although the finding of a genetic influence on objectively-assessed AL provides unique empirical support for theories of temperament that include AL as a dimension and presume a constitutional or biological foundation, it contributes little information about genetic influences on AL from a developmental standpoint. Because genetic influences are not necessarily stable across the lifespan, the finding of a genetic effect on AL at one age, in this case, infancy, tells nothing about such influences at another age. It may be the case that the importance of genetic influences on AL vary as a function of age. Therefore, any attempt to generalize these results to another age level would be presumptuous.

Genetic Influences on AL: A Developmental Perspective

Fundamental Questions

For the developmentalist, the major marker of genetic and environmental change is age (Plomin et al., 1988). Consequently, age is a variable that must be considered when examining the genetics of temperament. There are two fundamental questions that can be explored in developmental behavioral genetics. The first, inquires about differential heritability across ages. This issue is important because the investigation of the etiology of individual differences over age may serve to identify points of causal transition. The second question posed by the developmental behavior geneticist explores the role of genetic influences on the change or continuity of individual differences during development. Thus, this question addresses the process by which developmental change takes place. The questions asked by the behavior geneticist will guide the choice of developmental research design to be employed (i.e., cross-sectional versus longitudinal).

Differential Heritability Across Age

As a result of the dynamic nature of gene action throughout the lifespan, one area of interest in developmental behavioral genetics is the question of differential heritability of behavior as a function of age. It is no longer appropriate to presume that studies of genetic influences in one age group yield valid information with regard to other age groups (Dworkin, Burke, Maher, & Gottesman, 1976). The degree of heritability or, indeed, whether or not a trait is heritable, may vary across age spans.

Thus, research exploring the issue of differential heritability asks whether the mixture of genetic and environmental influences changes with development (Buss & Plomin, 1984). As the relative influences of genetic and environmental factors change, so too will the estimate of heritability.

Plomin et al., (1988) define heritability as a descriptive statistic that specifies the "portion of observed variability of a given trait that can be accounted for by genetic differences among individuals in a particular population at a particular time" (p. 113). Thus, a change in heritability would reflect a change in the proportion of individual differences (observed behavioral variation) that can be attributed to genetic variance. For example, as indicated earlier, in the twin study, doubling the difference between MZ and DZ intraclass correlations provides a quantitative estimate of heritability. An increase in heritability as a function of age would therefore, be the consequence of an increased difference between MZ and DZ intraclass correlations. This would transpire if MZ co-twin similarity increased to a greater extent than DZ co-twin similarity, or conversely, if the DZ co-twin similarity decreased to a greater extent than that of MZ co-twins (Plomin et al., 1988).

In developmental psychology there are relatively few studies exploring the issue of age-related changes in the magnitude of heritability. This is not surprising given the substantial statistical hurdle that must be overcome to answer questions of this nature. Power, a concern with most twin studies, is an even more severe problem when attempting to compare two heritabilities. According to Buss and Plomin (1984), with a sample of 100 pairs of each twin type, a heritability estimate of .40 will not differ significantly from zero. With a sample of 1,000 pairs each of MZ and DZ twins, it would be possible to make this discrimination, but not to discriminate between heritabilities of .40 and .20. Consequently, few studies can actually test the significance of the difference between heritabilities at different ages. Instead, speculations about the tendency for heritability to display developmental increases or decreases are made on the basis of the changing pattern of MZ and DZ intraclass correlations across age (e.g., Plomin et al., 1988). For this reason, the results regarding changes in heritability should only be viewed as suggestive; not conclusive.

Genetic Contributions to Developmental Change and Continuity

Another developmental approach that has resulted from the understanding of the dynamic nature of genes is the analysis of genetic and environmental contributions to change and continuity in development. Independent of changes in heritability across age, genetic influences can contribute to developmental change and continuity. Unlike the differential heritability approach, the focus in this second approach is on *the process* by which developmental change takes place (Plomin et al., 1992).

The contemporary question confronting behavioral genetic researchers therefore, becomes how much, if any, change or continuity of temperamental individuality can be attributed to genetic influences (Matheny, 1983). A first step in answering this question involves establishing the degree to which individual differences are constant across ages. After evaluating ontogenetic change and continuity, one can investigate its sources.

Individual differences and development. Longitudinal consistency for a behavior is usually indexed by the relative constancy of individual ordinal position within a group (i.e., cross-age correlations). McCall (1977) criticizes developmental research for its emphasis on stability, not change, in individual differences, and cites the use of cross-age correlations as an example. However, this statistic can also serve as a gauge of developmental discontinuity, the reordering of individual differences across age. Age-to-age correlations that are less than the reliable variance of the measure reflect genuine developmental change (Clarke & Clarke, 1984). Thus, when evaluating stability or change in individual differences, it is critical to know the immediate test-retest reliability of the measure in order to separate measurement error from developmental change. Unfortunately, this comparison is rarely made in developmental literature.

Assessing the role of genes in developmental change. Wilson (1986) notes that genetic factors have traditionally been considered to promote ontogenetic continuity. Following this logic, it was assumed that age-to-age stability was required to infer a genetic effect. Age-to-age changes in individual differences were, thus, interpreted as evidence against genetic influences. This position, however, ignored the possibility "that change itself might be a reflection of systematic genetic influences on development" (Wilson, 1986, p.48). Discontinuity in behavioral development may indicate a pre-programmed change in the activation of genes (Wilson, 1983). If change in individuals is due in part to timed genetic (i.e., chronogenetic) influences, in a twin study, MZ twins should show significantly greater within-pair similarity for developmental change than DZ twins. Accordingly, a pattern of low to moderate behavioral stability and significantly greater MZ concordance for change would suggest that the reordering of individual differences of temperament is partially regulated by genetic influences. Because infancy is a period of rapid developmental change, Wilson (1981, 1983) observes that the study of developmental processes in infant twins is a "powerful resource" for investigating the role of genetic influences in guiding behavioral development.

Assessing the role of genes in developmental continuity. In addition to evaluating genetic influences on individual differences in mean level change, one can assess genetic contributions to phenotypic stability across age. The finding of a significant genetic influence on phenotypic stability implies that there is some overlap between the genetic factors that affect the trait at the two ages (Plomin, 1986a). In the twin study, the role of genetic influences on developmental continuity is assessed via cross-twin intraclass correlations. That is, the twin's score at Time 1 is correlated with the other twin's score at Time 2. If phenotypic stability is mediated by genetic factors, then the cross-correlations for MZ twins should be greater than the DZ twin cross-correlation (Plomin & Nesselroade, 1990). The logic behind this is quite simple. Significant differences between MZ and DZ intraclass correlations for a single occasion suggests heritability at that moment. A simple two-occasion (lagged) correlation is a typical measure of phenotypic stability. Thus, a significant difference between MZ and DZ cross-twin intraclass lagged correlations suggests a genetic contribution to phenotypic stability in individual differences. Just as doubling the difference between single occasion MZ and DZ intraclass correlations estimates heritability, doubling the difference between MZ and DZ cross-twin correlations estimates the "heritability of stability" (Plomin, 1986b).

As Plomin et al. (1992) point out, although they are related, genetic change and genetic continuity are not simply different sides of the same developmental coin. That is, while there may be genetic influences contributing to *both* developmental change and continuity for a given trait, it is also possible that genetic factors influence change, but not continuity, or vice versa.

Longitudinal Versus Cross-sectional Designs

In deciding whether to conduct a longitudinal or cross-sectional study a researcher must distinguish whether the goal of the research is to explore age *differences* or age *changes* (Wohlwill, 1973). According to McCall (1977), if the research seeks to describe how children at various age groups differ, then cross-sectional research methods are appropriate. However, the exploration of developmental change within individuals and the factors that promote development dictates that a longitudinal study is required. For this reason, the longitudinal methodology is the lifeblood of developmental psychology (McCall, 1977).

From a developmental behavioral genetics perspective, these distinctions relate directly to the questions of differential heritability across age and the role of genetic influences in developmental change. Cross-sectional research can only address the question of whether the heritability of a trait differs at one age versus another. In contrast, the longitudinal design can evaluate phenotypic stability, detect the presence of genetic influences at two or more periods of development, and investigate the possibility of genetic influences operating on ontogenetic change and continuity within individuals. For these reasons, the longitudinal design is, generally, the preferred method in developmental behavioral genetics.

Empirical Findings

Although there is an abundance of behavioral genetic research examining the etiology of individual differences in AL and other temperamental dimensions, there is a relative dearth of research exploring this issue from a developmental perspective. At present, the few studies that do take this approach are predominantly twin studies.

Cross-sectional Studies

Despite the aforementioned limitations of the cross-sectional design, there have been a number of studies addressing the question of differential heritability of activity level across age (e.g., Buss et al., 1973; Matheny, Dolan and Wilson, 1976; Stevenson & Fielding, 1985; Willerman, 1973). Overall, the data are equivocal, with results varying according to the measure employed and the specificity of the judgments required in rating activity level.

Willerman (1973) had mothers rate the activity of their twins with the Werry activity level questionnaire. This rating scale involves judgments about specific behaviors in specific situations. The subjects in this study ranged in age from 11 months to 156 months. To examine age differences in heritability, Willerman split the distribution at the group's mean age of 50 months. Within both age groups, the MZ intraclass correlations were significantly larger than those of the DZ twins, with no apparent contrast or assimilation effects. Overall, there was no evidence to suggest that genetic influences varied across age groups. The scale displayed similar patterns of MZ and DZ co-twin similarity regardless of the age of the twins (e.g., above 50 months R_{MZ} = .89, R_{DZ} = .52; below 50 months R_{MZ} = .92, R_{DZ} = .62).

In a similar approach, 127 mothers in Buss et al.'s (1973) twin study made global judgments regarding their children's temperament on the EASI questionnaire. The subjects in this investigation ranged in age from 4 months to 16 years. As in Willerman's (1973) study, two age groups were formed by dividing at the mean age (55 months). The age groups were further divided by sex, making the number of twins within each group small. Thus, the power to detect significant differences between groups was quite low, limiting any conclusions that might be made on the basis of null findings. The results in this study were inconsistent. Across both ages, male MZ twins were significantly more similar than male DZ twins, whereas for females, the difference between MZ and DZ similarity was significant for only the older age group (e.g., males under 55 months R_{MZ} = .91, R_{DZ} = .47; males over 55 months R_{MZ} = .73, R_{DZ} = .00; females under 55 months R_{MZ} = .68, R_{DZ} = .58; females over 55 months R_{MZ} = .70, R_{DZ} = .00). For both sexes, the DZ twins displayed moderate intraclass correlations at the younger age level, but those in the older group had intraclass correlations of zero. On the basis of these results, estimates of heritability increase with age. However, the differences between MZ and DZ correlations in the older age group are much larger than would be predicted from the classical twin model and suggest the presence of contrast effects.

It should be noted that the analysis of differential heritability was not the primary goal of either of these studies. Consequently, the formation of age groups was rather arbitrary. An obvious problem with this approach is that the constructed age groupings are not intrinsically meaningful. Further, there remains a wide range of ages within each group and hence, some developmental differences in heritability could go undetected.

Differential heritability across age was a focus in Stevenson and Fielding's (1985) study using the EASI to examine the temperamental similarities of over 200 pairs of twins in three age groups; 0-2 years, 2-5 years and over 5 years of age. This large study yielded results comparable to Buss et al. (1973). At all age levels, MZ twins were substantially more similar for AL than DZ twins. The magnitude of MZ-DZ differences increased with age, suggesting increasing genetic influences from infancy to middle childhood. However, the pattern of resemblance once again, violates the twin model. Overall, the DZ correlations were extremely low and became increasingly lower (i.e., negative) with age. Thus, the apparently increasing heritability might be a reflection of increasing contrast effects.

More recently, a large twin study (*N* = 306 twin pairs) conducted by Cyphers, Phillips, Fulker and Mrazek (1990), failed to find differential heritability on parent-rated activity level from infancy to early childhood (i.e., twins ranged in age from 6 months to 4 years). In this study, parents rated their twins' temperament on either Carey's Infant Temperament Questionnaire (ITQ) or the Toddler Temperament Questionnaire (TTQ). Using multiple regression analyses predicting a twin's score from it's co-twin's activity score, gender, age, and the product of age and co-twin's activity score, the authors found no significant age or gender effects for either MZ or DZ twin groups. Therefore, for this sample, twin similarity on parent-rated activity did not significantly vary as a function of age. Pooling the data across age and gender, genetic influences were estimated to account for 57% of the observed individual differences in activity.

In one of the few studies that do not employ parental ratings of temperament, Matheny, Dolan and Wilson (1976) had observers rate the behavior of twins with the Infant Behavior Record (IBR) from the *Bayley Scales of Infant Development* [BSID] (Bayley, 1969). The IBR is a global rating scale which comprises items representing broad dimensions of infant behavior including activity level. It is typically completed by the observer within the context of a specific, somewhat stressful situation (i.e., during administration of the BSID). In the Matheny et al. study, the sample comprised two small, longitudinal subsamples. The first had been observed at 3-, 6-, 9-, and 12-months of age; the second at 18-, 24-, and 30-months of age. First and second year summary activity scores were obtained by totaling the activity ratings from each observation period. For each age grouping, the intraclass correlations of the MZ twins was significantly greater than that of the DZ twins, suggesting genetic influences on activity. Although the pattern of MZ-DZ twin similarity for activity level was not substantially different across age, energy, a related factor showed evidence of a genetic influence only during the second year.

Longitudinal Studies

Despite the wealth of information a longitudinal design can offer, as the following review will show, few developmental behavioral studies have exploited its full potential. Until recently, the search for age changes in the heritability of temperament dimensions has been the most frequent research objective. In contrast, evaluation of the role of genetic influences on developmental change has been rare.

Employing a standardized version of the NYLS parental interview method, Torgersen and Kringlen (1978) found that the activity category displayed no evidence of significant genetic variance at 2 months of age (i.e., no significant difference between the within-pair variances of MZ or DZ twins); however, at 9 months, significant genetic variance emerged as indicated by within-pair variances that were significantly greater for DZ twins. Across age, the average mean differences between MZ co-twins activity remained relatively constant, whereas DZ mean differences for activity increased substantially. Significant genetic variance for AL was also found in a follow-up study conducted when the subjects were 6 years old (Torgersen, 1981). With age, the *F* ratios of DZ within-pair variance to MZ within-pair variance became increasingly larger, prompting Torgersen to conclude that the genetic component of activity level is not fully expressed at birth but becomes apparent and stronger as the child develops. However, twin intraclass correlations, calculated only at the 6 year age level, are a cause for concern. The problem of a DZ correlation that is much less than half the MZ correlation once again suggests that contrast effects may be operating to make DZ co-twins less similar (e.g., $R_{MZ} = .93$, $R_{DZ} = .14$).

Goldsmith and Gottesman (1981) analyzed temperament data in a sample of approximately 350 twin-pairs participating in the National Collaborative Perinatal Project. In this study, psychologists rated child temperament on scales similar to the IBR. Ratings were available for ages 8 months, 4 years and 7 years. During infancy, co-twin similarity comparisons indicated a genetic component for AL. At age 4, no substantial genetic influences were apparent. Although at age 7 there was a significant difference between MZ and DZ correlations, the problem of lower than expected DZ co-twin similarity once again emerges. Overall, intraclass correlations for MZ twins showed no change over age, whereas, DZ intraclass correlations fluctuated considerably across age levels. Although these results may reflect age dependent variations in AL, the general inconsistency of these results may also be due to variations in temperament structure across age. The cross-age comparisons were of similar, but not necessarily equivalent dimensions, and Goldsmith and Gottesman note that at the 4 year age level, the only age not to show a genetic influence, the activity factor was more than a quantitative measure of motor activity.

The Louisville Twin Study (LTS) is one of the most extensive ongoing longitudinal twin studies of infant and child behavior. Although the initial focus of this project was on mental abilities, in recent years it has made many contributions to the field of temperament. The emphasis of the LTS research program is on evaluating the structure of temperament across age, the stability of temperament measures over age, and the appearance of synchronized developmental trends for twins (Wilson & Matheny, 1986).

In an early LTS study of temperament, Matheny, Wilson, Dolan and Krantz (1981), conducted a series of parent interviews over the course of twins' first six years of life. Mothers were asked to judge their infant twins as the same or different for a variety of behaviors. Though this technique yields strictly relative, qualitative data and is not amenable to parametric analysis, the authors note that it is a practical approach to highlighting differences within twin pairs. With regard to activity, there were significant age-to-age associations from 6 to 24 months, 24 to 48 months and 36 to 72 months. Thus, the ascription of similarity at one age level was significantly associated with the same ascription at the later age level. Activity level typically provided sharp contrasts between twin types. At 6-, 24-, 36-, 48-, and 72-months of age, the concordance rates (expressed in terms of percentages) for MZ co-twins were consistently higher than for DZ co-twins. These results were significant for 4 out of the 5 age levels, only at 48 months of age did the differences in MZ and DZ concordance rates emerge as nonsignificant. Hence, this study suggests that activity level continues to display genetic influences from 6 months to 6 years of age.

In addition to parent ratings, the LTS has also used rater observation with the IBR as a measure of infant behavior. In a mixed longitudinal design, Matheny (1980) factor analyzed the IBR in a sample of infant twins at 3-, 6-, 9-, 12-, 18-, and 24-months of age. An activity factor was one of three recurrent factors to be found at every age. Twin intraclass correlations were calculated at each of these age levels (see Table 1). As can be seen in Table 1, the difference between MZ and DZ co-twin

Table 1. Twin Intraclass Correlations for Activity Factor Adapted From Matheny (1980)

Age in months	R_{MZ}	R_{DZ}	R_{MZ} - R_{DZ}
3	.30	.33	-.03
6	.24	.11	.13
9	.25	.22	.03
12	.33	.28	.05
18	.43	.14	.29*
24	.53	.14	.44*

Note. * $p \leq .05$, $R_{MZ} > R_{DZ}$.

similarity was negligible at 3 months of age; however, with progression in age, the contrasts becomes more pronounced and reaches significance during the second year.

The results suggest that activity demonstrates an apparent trend towards increasing heritability with age. However, Matheny notes that the absence of genetic influences during the first year might be due to the fact that at the earlier ages the activity factor had additional loadings from scales that have not shown evidence of genetic influences.

Matheny (1983) has demonstrated significant age-to-age stability for the IBR factors at 6-, 12-, 18-, and 24- months of age. As is characteristic of longitudinal psychological research, the age-to-age correlation matrix from this study was simplex in form, that is, the greater the age span between assessments the lower the correlation (Clarke & Clarke, 1984). Generally, the correlations between adjacent age levels were low-to-moderate, becoming higher in the second year (e.g., 6-12 months r = .29; 12-18 months r =.34; 18-24 months r = .42). This pattern of moderate stability would suggest that some developmental change had taken place. An analysis of twin concordance for change found that during the ages from 12 to 24 months, within-pair similarities (i.e., intraclass correlations) for change profiles on the activity factor were significantly higher for the MZ twins than for the DZ twins (e.g., $R_{pMZ} = .52$, $R_{pDZ} = .18$). Thus, it would appear that the pattern of ontological change in the second year of life is, in part, regulated by genetic influences.

More recently, the LTS has incorporated a laboratory assessment of temperament and the Toddler Temperament Questionnaire, a parent rating scale, into their research program (Wilson & Matheny, 1986). The stability of AL appears to vary according to the measure employed. The age-to-age correlations for videotaped laboratory observations of activity level were moderate and generally similar for the 9-to-12, 12-to-18 and 18-to-24 month intervals. In comparison, stability correlations for the parent questionnaire rating of activity were considerably higher and increased across these age intervals. Because their study is preliminary in nature, genetic analyses were conducted only for the laboratory measure. The pattern of twin intraclass correlations at 9-, 12-, 18-, and 24-months of age suggest a genetic influence on AL during the second year. However, due to the small sample size, the evidence is merely suggestive; only at 18 months of age were MZ and DZ correlations significantly different. Moreover, this apparent change in the heritability of AL might be measurement artifact. In a factor analysis of laboratory ratings across age, activity loadings declined at 18 and 24 months, suggesting that a change in the organization of activity level is contemporaneous with the increase in heritability.

The most comprehensive twin study exploring genetic change and continuity is the MacArthur Longitudinal Twin Study [MALTS] (Plomin et al., 1992). The MALTS is a large, collaborative research project focusing on individual differences in change and continuity in temperament, emotion, and cognition from infancy to early childhood (Plomin et al., 1990). The developmental behavior genetic analyses in this study capitalize on the rich information available in longitudinal research by exploring phenotypic stability; differential heritability; and genetic contributions to both, ontological change and phenotypic stability.

Using a multimethod approach, in the MALTS, temperament is assessed via parent ratings on the CCTI, and examiner ratings of infant and child behavior on the IBR. These measures yield different results in an analysis of genetic continuity and change of activity level for 200 pairs of twins assessed at 14 and 20 months of age (Plomin et al., 1992). Although the cross-age stability correlations for activity level are significant for both parent and examiner ratings, when assessed via parent ratings, AL demonstrated higher phenotypic stability (r_{parent} = .64, $r_{examiner}$ = .24). The heritability of activity was significant at 14 and 20 months for both measures; however, at both ages, the DZ intraclass correlation for the CCTI was negative, suggesting a possible contrast bias. Overall, no evidence for differential heritability of activity emerged from either the CCTI or IBR ratings.

An evaluation of the etiology of age-to-age change in activity revealed significant genetic influences on CCTI change scores, but not for IBR change scores. Thus, only for parent ratings of activity did MZ co-twins display significantly greater within-pair similarity for developmental change than DZ co-twins. The etiology of phenotypic stability displayed a different pattern of results. For both measures, MZ and DZ cross-twin correlations were low and did not yield significant genetic influences on the phenotypic stability of activity. However, the estimate of IBR heritability from this cross-twin stability analysis (h^2 = .22) was close to IBR phenotypic stability. This suggests that genetic factors account for nearly all the phenotypic stability across the ages studied, and that the failure to find significant heritability is a consequence of low statistical power. Longitudinal model-fitting analyses confirmed the findings of a significant genetic change for CCTI activity, and the genetic mediation of IBR phenotypic stability. Moreover, the analyses indicate that nonshared environmental influences, possibly stable rater biases, contribute to the stability of CCTI activity.

At present, the Colorado Adoption Project (CAP) is the only non-twin study to examine genetic influences on infant and child temperament from a developmental perspective (Plomin, et al., 1988; Plomin, et al., 1991). The CAP is a longitudinal, full adoption design which is currently exploring the origins of individual differences in infancy and early childhood. Within this design, genetic influences are evaluated through parent-offspring and sibling adoption analyses.

In the CAP, temperament was assessed by parent ratings on the CCTI and rater observation on the IBR (Plomin & DeFries, 1985; Plomin et al., 1988). As was apparent in the LTS and MALTS data, year-to-year stability for AL in the first four years of life shows varying results depending on the method of assessment. Parent ratings of temperament on the CCTI displayed a pattern of moderate stability that increases across the first two years. In contrast, stability correlations for observer ratings of AL on the IBR are lower and relatively consistent across years (Plomin et al., 1988).

In an analyses of genetic influences on parental ratings of temperament in infancy and early childhood, Plomin et al. (1991) examined parent-offspring correlations for temperament in three parent-offspring relationships: biological parents and their adopted away offspring, nonadoptive parents and their nonadoptive offspring, and adoptive parents and their adopted offspring. Adult temperament assessed on the EASI, was then correlated with child temperament ratings from 1 to 7 years of age. The data revealed no apparent age trends, nor was there evidence of significant genetic effects.

The sibling adoption analysis of CCTI data for ages 1 to 4 yielded similar results. With this design, genetic influences are suggested when adoptive sibling correlations are significantly lower than those for nonadoptive siblings. For the CAP parent rating data, this comparison revealed no consistent pattern of results; both adoptive and nonadoptive sibling pairs demonstrated low intraclass correlations. According to Plomin et al., the difference between adoption and twin results suggest that, as a result of contrast effects, assimilation effects, or nonadditive genetic variance, twin studies may exaggerate the magnitude of genetic influence on AL and other temperament dimensions.

In contrast to parent-rated temperament, CAP sibling adoption data for observer ratings of temperament during infancy does, however, suggest genetic influences on activity level. Braungart, Plomin, DeFries and Fulker (1992) compared CAP sibling adoption data with twin data from the LTS (Matheny, 1980) for IBR ratings of temperament at 1 and 2 years of age. At both ages, the intraclass correlations for IBR activity produced a pattern suggestive of genetic influence (i.e.,

$R_{MZ} > R_{DZ} = R_{Nonadoptive} > R_{Adoptive}$). Moreover, there was some suggestion of an increase in genetic influence on AL as evidenced by MZ correlations increasing to a greater extent than DZ correlations. Using maximum-likelihood model fitting methods to compare the data from the four groups simultaneously, IBR ratings of activity evinced significant genetic influences at 12 and 24 months of age. Although once again, the data was suggestive of an age-related increase in heritability, the effect was not statistically significant. Thus, averaging across the two ages, heritability accounts for approximately 47% of the variance in infant IBR activity, with the remaining variance attributed to nonshared environment. In addition, the model-fitting analyses indicated that twin and sibling adoption data do not yield significantly different results.

The discrepancy between parent and observer ratings of activity level in the CAP suggest a possible rating bias for the CCTI. Parents in both the adoptive and nonadoptive groups were required to rate the AL of two children, although parents did not rate each child's AL at the same time. Nonetheless, parents may well have rated the second child in the context of the first child's behavior at the same age. Thus, ratings would reflect parental expectations as a result of contrasting the siblings and highlighting behavioral differences. This notion is supported by the presence of negative correlations for temperament variables between both biological and adoptive siblings.

Conclusions

Although the studies reviewed clearly demonstrate the importance of genetic influences on infant and child activity level, there are no consistent developmental trends across studies. The research examined encompasses a wide range of ages and a variety of temperament measures, and this likely obfuscates the developmental picture. Nevertheless, some general comments can be made.

A primary focus of these investigations has been on whether or not there is a change in the relative influences on genetic and environmental influences with development. Intuitively, one might posit that as a child matures and becomes more interactive with increasingly diverse environments, the role of genetic factors might wane. The research examined does not support this notion. Although the results were somewhat variable, there was no trend towards decreasing genetic influences. Heritability, as indicated by the pattern of MZ and DZ intraclass correlations, either remained constant or increased with age.

Based on the reported research, conclusions about the differential heritability of AL across age must be made cautiously. As indicated earlier, due to the huge sample sizes required, it is impossible, from a practical standpoint, to test for significant differences between heritability estimates. Consequently, in most twin studies, inferences about changing heritability are made on the basis of relative changes in the similarities of MZ and DZ twins. Often, an increase in genetic influence is inferred when the difference between MZ and DZ intraclass correlations is nonsignificant at an early age and significant at a later age. While this approach seems logical, it fails to take into consideration the limitations of low statistical power. Several of the studies reviewed had relatively small twin samples (i.e., less than 100 twin pairs). In these cases, it is imprudent to interpret a nonsignificant finding as evidence for no genetic influence. The studies that had large samples (i.e., Cyphers et al., 1990; Goldsmith & Gottesman, 1981; Plomin et al., 1992; Stevenson & Fielding, 1985), produced conflicting evidence with regard to differential heritability of AL across age.

Assuming that power was not a problem, a number of explanations for a trend towards increasing genetic influences on AL can be entertained. First, genetic influences on AL in early infancy might be masked because of perinatal environmental influence (Torgersen, 1985). This would appear to be the case for physical development (Wilson, 1979a) and may well hold for the development of temperament. Second, Plomin and DeFries' (1985) "amplification model" might apply. According to this developmental model, genetic effects that create small individual differences in infancy become magnified with age. Third, Scarr and McCartney (1983) posit a shift from passive to reactive and active genotype-environment correlations as an explanatory mechanism for why DZ twins become increasingly different with development. The similarity of DZ twins' early environments, which are passively correlated with their genotypes, give way as they actively select environments correlated with their different genotypes. Because MZ twins select highly correlated environments, this shift from passive to active genotype-environment correlations would not decrease their similarity. However, because, in the studies reviewed, the pattern of increasing heritability across age was not ubiquitous, a more plausible explanation might lie within the measures used to assess AL.

Temperament measures may have differential reliability or validity across age. The prevailing research examining the genetics of AL employ subjective, parent or rater observations as a standard system of measurement (Aftanas, 1988). Unreliability of the ratings would tend to diminish the likelihood of finding significant genetic effects (Goldsmith & Gottesman, 1981). For temperament and personality, the measurement properties of the assessment techniques, especially during infancy, are often weak and are apt to systematically affect the probability of finding genetic influences (Goldsmith, 1983). Buss and Plomin (1984) suggest that during infancy there may not be enough behavior to obtain a good assessment of AL. While this, may or may not be the case (cf. Eaton, McKeen and Lam, 1988) it may be harder for judges to differentiate behavior in infants. If the reliability of the rating measure improves from infancy to early childhood, then the probability of finding a genetic influence would improve concomitantly. Although for temperament rating measures, the study of differential reliability across age has, for the most part, been neglected, it is frequently acknowledged that measures may have lower reliability at younger ages (e.g., Goldsmith, 1983; McDevitt, 1986; Torgersen, 1985).

Similarly, differential measurement validity across age could produce a pattern of variable heritability. Rater bias might increase with age. Neale and Stevenson

(1989) postulated rater bias as a possible cause of contrast effects on the EASI parent rating scale of temperament. It is interesting to note that, in the studies employing parent ratings, those that suggested increasing genetic influences also showed a trend towards increasing contrast effects (e.g., Buss et al., 1973; Stevenson & Fielding, 1985; Torgersen, 1981; Torgersen & Kringlen, 1978). Contrast effects result from the rater's tendency to contrast one twin with the other thereby magnifying their behavioral differences in the process. If this bias increased with age, the DZ intraclass correlations would become progressively lower and the probability of a significant difference between MZ and DZ correlations would rise. Additionally, apparent changes in heritability for AL might reflect differences in item content or in the factor structure of the measure at different ages. Indeed, when differential heritability was suggested in studies employing observer ratings, there were concurrent changes in factor structure (e.g., Goldsmith & Gottesman, 1981; Matheny, 1983; Wilson & Matheny, 1986).

Differential measurement reliability and validity would effect the phenotypic stability of AL in a similar manner. The studies reviewed suggested some stability in AL across ages. This is consistent with previous research and conforms with temperament theory (McDevitt, 1986). However, the trend towards increasing stability with age, although common in the temperament literature, might be an artifact of the measures employed. As with heritability, increases in the reliability or validity of the measure would result in apparent increases in stability of AL across age. In addition, the stability estimates might be reflecting the stability of rater behavior rather than stability in AL itself. As compared to observer ratings, parent ratings yield higher estimates of stability, which in some cases, become increasingly larger with age (e.g., Plomin et al., 1988; Plomin et al., 1992; Wilson & Matheny, 1986). This difference suggests that to some extent, it is the parent perceptions of AL and rating behaviors that are stable.

In contrast to the question of differential heritability, the question of genetic influences on developmental change and stability has received little attention in the current literature. Extant studies yield ambiguous results. In the MALTS (Plomin et al., 1992), genetic influences on change in AL from 14 to 20 months were significant for parent rated AL, but not for IBR measures of AL. However, in the LTS (Matheny, 1983) change in IBR activity from 12 to 24 months was significantly heritable. The different age spans studied (6 versus 12 months) might account for the discrepant IBR outcomes. That is, because the LTS spans 12 months, there might be more reliable change variance to capture, resulting in a more powerful test of genetic influence. While this might be the case, it is apparent that the finding of genetic influences on developmental change in AL from infancy to early childhood must be replicated before any firm conclusions can be drawn.

Similarly, the question of genetic contributions to phenotypic stability has not been clearly answered. Parent ratings suggest that stability in AL is environmentally mediated, whereas observer ratings for the same sample display evidence of genetic influences on continuity in AL (Plomin et al., 1992). Stable rater biases, such as contrast biases, are a potential environmental (i.e., nongenetic) source of phenotypic stability in parent ratings. Certainly, there is evidence to suggest that the stability of AL found with parent ratings might, to some extent, reflect the stability of parent perceptions. However, it must also be considered that parent and observer ratings sample different behaviors, and this could account for the contradiction in results.

The possibility that the developmental trend towards increasing heritability and stability of AL is an artifact of the differential accuracy of the subjective rating measures highlights the need for a more objective approach. Although they provide a more objective measure of temperament than do parent ratings, observer ratings are narrow in scope and permit only limited behavioral sampling. More extensive sampling is now possible, however, through the use of mechanical instruments to record AL, and this measurement option addresses the critical issues raised above. Mechanical devices eliminate the need to involve a person to observe and evaluate AL. Thus, they are a much more objective method for assessing behavior; permitting the transition away from human inferences based on qualitative behavioral definitions to a more quantitative measure of the physical forces that are associated with human activity. Mechanical instruments should not demonstrate differential accuracy across age because they simply record *all* occurrences of activity above the device's threshold level. Thus, the potential confounding influences of the stability of rater behavior and perceptions on the stability of AL does not become an issue when evaluating activity with physical instruments. In addition, mechanical measures of AL share the strengths of parent ratings, namely, their ecological validity and cross-situational generality, while circumventing the problems of rater bias. Despite these advantages of mechanical measures of AL, they have not been employed to answer developmental behavior genetic questions about temperament.

The Present Study

There is, clearly, a need for a developmental analysis of genetic influences on AL that is not clouded by the subjective errors that may arise when human judgment is employed as the measurement standard system. To address this, the present research longitudinally extended an initial, twin study of infant AL by re-evaluating the influence genetic of factors on activity level during early childhood (i.e., between 2½ and 3 years of age). The longitudinal design permitted the evaluation of both the question of changes in the presence of genetic influence across age, and the question of genetic influences on the continuity and change of individual differences in AL during this developmental period.

An additional issue not addressed in previous twin studies of infant and child temperament is that of the potential confounding influences of maturational factors. With infants, AL has been found to show a positive relationship with motor development (Escalona, 1968; Fagan, Singer, Ohr, & Fleckenstein, 1987; Fish & Crockenberg, 1981; Matheny & Brown, 1971). Furthermore, motor development has maturational underpinnings, and if motor development is a foundation for AL, the heritability of AL will be a result of the heritability in maturational timing. Thus, it is plausible that MZ twins could be more similar in AL than DZ twins simply because they are more similar in level of motor maturity.

In our initial twin study, motor development, as assessed on the Bayley Scales of Infant Development (Bayley, 1969), yielded evidence of a significant genetic influence and confirmed other published reports (e.g., Wilson & Harpring, 1972). In addition, the presence of a high DZ correlation in our infant sample implies that shared environmental influences were also operating. The hypothesis that concordance in motor development could account for concordance in activity level was considered, and though the Bayley Motor Scale was significantly associated with AL, MZ and DZ co-twin similarities were little altered following adjustments for motor maturity differences.

Although in infancy, similarity in AL is not an artifact of similarity in motor maturity, the question remains as to whether this will be the case for early childhood. The transition from infancy to early childhood is marked by the emergence of more efficient and skilled motor behaviors which may unmask individual differences that had been obscured by motor immaturity. Thus, the relation between motor development and activity level may become more salient. Additionally, it is possible that increases in the heritability of motor maturity could account for apparent increases in the heritability of activity level.

Hypotheses

Based on the findings of the initial twin study, the empirical studies reviewed, and the issues previously discussed, a number of hypotheses can be put forth:

1. For the sample of subjects participating in both phases of the research project, it was predicted that activity level, as assessed by actometers, would show evidence of a genetic influence at both ages (i.e., during infancy *and* early childhood). Although there is a marked change in the child's interactions with his or her environment across this developmental period, the research reviewed consistently indicates a genetic influence on AL after the second year of life. It was expected that this would also be the case when AL is objectively assessed. Unfortunately, power restrictions precluded the evaluation of whether genetic influences on AL significantly differ over the two age levels.

2. Because objective measures of activity level were employed, it was anticipated that there would be no evidence of contrast effects (i.e., the problem of too low DZ co-twin similarity) at either age level. Thus, the pattern of co-twin similarity should conform to the classical twin model which proposes that DZ co-twin resemblance should be approximately one-half that of MZ co-twins.

3. Across the two ages, a reordering of individual differences in mechanically-assessed AL was expected. That is, phenotypic stability would be moderately low, indicating that, from an individual differences perspective, developmental change has taken place. Stability across time and situations is a common criterion for

temperament dimensions, and activity level is the single dimension to consistently evince modest age-to-age stability (e.g., Hubert et al., 1982; Matheny, 1983; Rothbart, 1981, 1986). However, because in the present study, the time interval between assessments was substantial, the law of simplex patterns of longitudinal correlations (Clarke & Clarke, 1984) lead to the prediction the age-to-age correlation was likely to be quite low.

4. Low age-to-age phenotypic stability in combination with significant genetic influences across age suggests that genes are a likely source of developmental change (Plomin et al., 1988). Consequently, hypotheses 1 and 3 together lead to the further prediction that developmental changes in activity level would show a genetic influence. That is, as found by Matheny (1983) and Plomin et al., (1992), MZ co-twins should show a greater concordance for *change* in AL than DZ co-twins.

5. Because it was predicted that AL would demonstrate low age-to-age stability, it was also predicted that cross-twin intraclass correlations for both MZ and DZ twins would be low. In addition, as a consequence of low statistical power arising from the low cross-twin correlations and the small sample size, it was expected that no significant genetic influence on phenotypic stability would be detected.

6. The long interval between assessments and the use of different instruments to assess motor maturity lead to the prediction that infant motor development would be a poor predictor of later motor development. Nonetheless, as was the case in the earlier study, it was anticipated that in early childhood motor development will be

positively related to activity level. Further, it was expected that motor development, because of its link with physical maturity, will show a genetic influence. However, although there may be a relationship between AL and motor maturity during early childhood, AL is not solely an epiphenomenon of maturational factors, but represents, at least in part, a maturation-independent individual difference dimension. Thus, after controlling for the influences of motor development, it was expected that MZ twins would continue to evince greater concordance in AL than DZ twins.

METHOD

Participants

Recruitment

Forty-six families who had participated in the initial phase of research were sent a letter describing the follow-up study, outlining its purpose and the nature of participant involvement (Appendix A). Approximately one week after the letter was mailed, parents were contacted by telephone to answer any questions regarding the study, and to ascertain their willingness to, once again, participate (see Appendix B for the telephone protocol).

Of the 46 families sent the initial recruitment letter, five could not be contacted by telephone (four had moved and could not be located, one did not respond to telephone messages or a follow-up letter). All 41 families who were contacted by telephone agreed to participate in the second phase of research. Thus, the overall acceptance rate including families both contacted and not contacted was 89%. One twin pair was dropped from analyses because it was an outlier on the distributions for chronological age (CA) and gestationally adjusted age (GA) at both the initial and follow-up assessments. A final sample of 40 twin pairs resulted.

Sample

The sample comprised 28 MZ twin pairs (12 female, 16 male) and 12 same-sex DZ twin pairs (9 female, 3 male), and was predominantly middle class according to the Hollingshead (1975) index of socioeconomic status (SES). The mean chronological age at the time of the first assessment was 7.3 months (SD = 1.2). Adjusted for prematurity, the mean gestationally-adjusted age, as calculated by the difference between the assessment date and original due date, was 6.5 months (SD = 1.3). MZ and DZ groups did not significantly differ in either CA ($M_{MZ} = 7.5$, $SD_{MZ} = 1.2$; $M_{DZ} = 6.8$, $SD_{DZ} = 0.9$) or GA ($M_{MZ} = 6.6$, $SD_{MZ} = 1.5$; $M_{DZ} = 6.1$, $SD_{DZ} = 0.8$).

At the follow-up assessment, the mean CA was 35.8 months (SD = 1.6) and the mean GA was 35.0 (SD = 1.6). The MZ twins were significantly older chronologically ($M_{MZ} = 36.2$, $SD_{MZ} = 1.5$; $M_{DZ} = 35.0$, $SD_{DZ} = 1.5$; $p < .05$), but not gestationally ($M_{MZ} = 35.3$, $SD_{MZ} = 1.6$; $M_{DZ} = 34.3$, $SD_{DZ} = 1.5$). This latter null finding is important because differences in gestational age may result in inflated twin correlations (Thompson, Fulker, DeFries, & Plomin, 1988).

The average interval between the initial and follow-up assessments was 28.5 months (SD = 1.8). MZ and DZ groups did not significantly differ on this variable ($M_{MZ} = 28.7$, $SD_{MZ} = 1.9$; $M_{DZ} = 28.2$, $SD_{DZ} = 1.8$).

Analysis for Possible Selective Attrition

A critical feature of any longitudinal study is the analysis for possible selective attrition -- one must determine whether the longitudinal sample is representative of the initial sample. To evaluate this, longitudinal subjects were compared with those subjects who did not participate in the second phase of the study. For each variable of interest, *t* tests were conducted to ascertain whether there were significant differences between the means of the two groups at the initial testing. In those cases where families contributed two values to a single variable (e.g., Twin A weight and Twin B weight), the mean value was taken and the analyses conducted on the family mean.

The subjects lost to attrition were younger in chronological age (M = 6.6 months, SD = 0.3, $p < .01$), but not in gestationally-adjusted age (M = 6.0 months, SD = 1.8). They did not significantly differ from the longitudinal subjects on the measures of actometer-assessed AL, parent-rated AL, motor development, head circumference or ponderal index. However, lost subjects were significantly lighter (M = 6.7 kg, SD = 0.9, $p < .05$), and shorter (M = 64.2 cm, SD = 2.8, $p < .05$), than the longitudinal sample. Overall, it would appear that selective attrition was not a problem in the present study.

Procedure

The follow-up study attempted to parallel the procedures of the initial research project as closely as possible (see Saudino & Eaton, 1991 for a description of the initial assessment procedures), and changes were made only where necessitated by the difference in age levels.

Overview

As in the first phase, the procedure involved two visits, 48 hours apart, to the participant's home. During the initial home visit, formal consent for participation was obtained (Appendix C). Following this, the parent was interviewed with regard to family demographics and the general health of the twins (Appendix D), and a brief assessment of each toddler's motor development was conducted using the Motor Scale of the *McCarthy Scales of Children's Abilities* [MSCA] (McCarthy,1972). The actometers were demonstrated and attached to the limbs of each child. The parents were provided with oral and written instructions regarding the care and use of the actometers (Appendix E) and with record sheets (one for each child) for the logging of the time that the actometers were off limbs (Appendix F). The motion recorders remained on the children for 48-hours; however, parents were free to remove them at any time. Parents were encouraged to engage in normal activity and to maintain daily routines with their children during the two-day data collection period. In addition, they were asked to complete by the second visit a questionnaire concerning the twin's degree of physical similarity.

The second visit was scheduled for 48 hours following the attachment of the actometers. At this visit, actometers were removed and read; the record sheets collected; a researcher assessment of physical similarity conducted; and physical measures of height, weight and head circumference taken. Parents were then asked to complete questionnaires rating the activity of each twin.

Instruments

Actometer Measure

During the two-day data collection period, activity level was objectively assessed with the Kaulins and Willis Model 101 Motion Recorders (see Appendix G for a discussion of actometer measurement properties). A single actometer weighs approximately 13 g. Each twin was randomly assigned a set of four actometers. To differentiate actometer sets within pairs, the back of each watch was covered with white surgical tape, on which the child's name was printed. Over the sample, actometers within each set were counterbalanced across limbs. Within a set, actometer watch-faces were color-coded (black, yellow, blue, red) and keyed to the parent recording sheet indicating the color of watch that went on each limb. Thus, the parent was able to easily identify the watch set that belonged to each twin and, further, to match the instrument to the correct limb.

After demonstrating the actometers to the parent, the start time of each watch was recorded. Next, the actometers were attached, one per limb, by means of plastic wrist bands with snap fasteners that locked so that the instrument could only be removed by cutting the strap. Arm attachment, at the wrist, was on the dorsal aspect of the forearm proximal to the radialcarpal joint. Leg attachment, at the ankles, was superior to the lateral malleoli. For each twin, parents were asked to record, on the sheets provided, the times when the actometers were off for baths, etc. Because removal of the watches required cutting the wrist bands, parents were provided with extra wrist bands and shown how to reattach the actometers. Generally, parents chose not to remove the actometers, and the mean time that the actometers were worn was 47.5 hours (SD = 1.5).

At the end of the 48-hour period following the attachment of the actometers, the watches were removed and a final reading of each instrument taken. For each limb, the number of activity units (AU), the total elapsed time in actometer seconds, was converted to a rate per 30 minutes real time measure. This conversion adjusted for the time that each watch was off a limb. As in the initial study, this rate measure was positively skewed; hence, a base 10 log transformation was applied to produce a more normal distribution. A composite actometer score, designed to reflect overall motor activity, was then calculated from the mean of the four logged limb actometer scores (see Appendix H for a summary of derived scores). To estimate actometer reliability, Cronbach, Gleser, Nanda, and Rajaratnam's (1972) generalizability approach was applied to the Twins A (40) x limb (4) matrix of AU log values. Variance components for Twins A, limb, and error were then estimated. The estimated reliability of a single actometer reading was .71, and the reliability of the score created by aggregating 4 limb scores was .91. This analysis for Twins B yielded comparable results (r_1 = .66, r_4 = .89).

Motor Development Measure

At the initial home visit, after having developed rapport with the parent and twins, each child's motor maturity was individually assessed with the Motor Scale of the *McCarthy Scales of Children's Abilities* [MSCA] (McCarthy, 1972). During the first phase of this research, motor development was assessed with the *Bayley Scales of Infant Development* [BSID] (Bayley, 1969). Because both scales evaluate relative motor maturity on a variety of fine and gross motor tasks, the McCarthy Motor Scale is a reasonable successor to the Bayley Psychomotor Developmental Index [PDI] (see Appendix G for a description of these measures).

Each toddler's scaled score (total number of items passed relative to his or her age) provides an index of maturity. Two examiners simultaneously scored the Motor Scale for 13 toddlers. Interobserver reliability, as calculated by the Pearson product-moment correlation of the two sets of scaled scores, was .98 $p < .0001$.

Parent Activity Questionnaire

In the initial study, at the end of the two-day data collection period, parents were asked to rate the activity of each twin on Rothbart's Infant Behavior Questionnaire (IBQ) Activity subscale (Rothbart, 1981). Because the IBQ is designed to assess the temperament of infants up to only 12 months of age, it was not suitable for the follow-up study. Consequently, Goldsmith's (1987) Toddler Behavior Assessment Questionnaire (TBAQ), which derives from the IBQ, was used to provide an age-appropriate rating measure of activity level (see Appendix G for a discussion of IBQ and TBAQ measurement properties). Parents were asked to rate each twin's activity level in a variety of specific situations observed during the two-day period in which the actometers were worn. For example, "when playing inside, how often did your child run through the house?" (see Appendix I for complete scale). Reliability for this measure was estimated to be .51 using Cronbach's alpha coefficient.

Physical Measures

At the second home visit, each twin, wearing indoor clothing, was weighed on a portable digital scale. Height was measured with an anthropometer, and head circumference was measured with a metric measuring tape. Two readings of each measurement were taken, and the mean values were regarded as the true measurements. Reliability was then estimated from the correlations between the first and second measurements. The intercorrelations for weight, height, and head circumference were .99, .99, and .98 ($p < .0001$) respectively. Based on the mean length and weight measures, ponderal index (PI) was calculated for each child (see Scanlon, 1984, for the calculation of PI). PI is a weight-for-length ratio of the relative amount of soft-tissue mass, and children with a high PI have more subcutaneous fat than children with a low PI.

Zygosity

Although the twins were classified as either MZ or DZ in the initial study, a second zygosity diagnosis was undertaken to enhance accuracy. Despite empirical evidence demonstrating impressive validity and reliability, the physical similarity questionnaires employed with our infant sample had not been previously used with subjects under one year of age (see Appendix J for a discussion of issues relating to the diagnosis of zygosity). It is possible that the discrimination of physical features is more difficult prior to one year of age, and hence, misdiagnoses may have occurred. The present study allowed an evaluation of this possibility.

Zygosity diagnoses from the initial study that were based on birth and delivery information were regarded as true. That is, if twins had different blood types (n = 1) they were classified as DZ. Similarly, a monochorionic placenta (n = 13) was regarded as proof of monozygosity (Thompson, 1985). In those cases where zygosity was not determined through birth and delivery information, zygosity was diagnosed through physical similarity criteria following the same procedure as in the earlier study. The zygosity classification from the second assessment was regarded as true, and all twin analyses (i.e., infancy *and* early childhood) were conducted on the basis of these revised twin groupings.

Both the parents and researcher independently participated in this procedure. The parents completed a written questionnaire concerning general and specific physical similarities between the twins and instances of identity confusion (see Appendix K). This questionnaire included a combination of items from the Nichols and Bilbro (1966) and Cohen, Dibble, Grawe and Pollin (1973, 1975) zygosity questionnaires as well as additional items designed to aid in making intuitive judgments. For the researcher assessment of twin physical similarity, the physical features (e.g., facial appearance; hair color, texture, amount and growth pattern; eye color; skin complexion; teeth patterns; and earlobe patterns) of co-twins were compared and ratings of "not at all similar", "somewhat similar", and "exactly similar" were applied (see Appendix K). To remain blind to the parents' responses, the researcher did not examine the parent questionnaire until after these ratings has been made.

Diagnostic decision rules adapted from Nichols and Bilbro (1966) and Plomin and Rowe (1977) were then implemented (see Appendix L). The classification of each pair was based on information regarding identity confusion that came from the parent questionnaire and from physical similarity judgments and measurements made by the researcher. Items were organized in a logical hierarchy, and zygosity classification was automatically assigned once item information was entered into a computer program. Using this system, it was possible to classify all twin pairs as either MZ or DZ.

Reliability of Physical Similarity Judgments

Judgments of physical similarity were evaluated through the interobserver agreement between parent and researcher. For each judgment, Cohen's kappa (κ) was calculated to assess beyond-chance agreement (see Table 2). These values ranged from .37 to .79 with a median value of .50, suggesting reasonable reliability. Judgments of eye color and teeth patterns were found to be most reliable, whereas amount of body hair and facial appearance had the lowest interobserver agreement. Generally, as compared to examiner ratings of physical similarity, parents were more inclined to see differences between their twins.

Validity of Zygosity Diagnoses

Carter-Saltzman and Scarr (1977) suggest that when questionnaires or ratings are to replace blood typing analysis of zygosity, the researcher should cross-validate the method in the sample to be studied. To provide a measure of validity for the

Table 2. Parent-researcher Agreement for Judgments of Physical Similarity

Physical Feature	κ
Face	.38
Hair color	.64
Hair curliness	.50
Hair thickness	.45
Hair growth pattern	.50
Amount of body hair	.37
Eye color	.79
Complexion	.52
Ear lobe	.45
Teeth	.67

second diagnosis of zygosity, the parent questionnaire for all twin pairs, including those diagnosed via birth information, was scored according to the Cohen et al. (1973, 1975) discriminant function method. The coded responses from the 10 Cohen et al. questionnaire items were multiplied by their corresponding discriminant function raw score coefficients and summed. Twin-pairs with a total above 26.71, Cohen et al.'s cutoff, were classified as MZ, and those below this value were considered DZ. Using the kappa statistic to assess the agreement of the two methods of zygosity classification, we found them to be highly congruent, $\kappa = .87$. This concordance accounted for 87% of the beyond-chance agreement between the classifications, a good level of consensus using a stringent criterion.

Using the zygosity diagnoses for the follow-up study as the true zygosity classification, it was then possible to evaluate the accuracy of the diagnoses made in the initial study. Kappa for the beyond-chance agreement between the zygosity classifications in infancy and early childhood was .83. Three twin pairs classified as DZ in infancy were reclassified as MZ in the present study. For two of these pairs, a flaw in the initial computer programme resulted in co-twins being treated as differing in blood type when, for each pair, there was information on the blood type of only one twin. Despite this, it would appear that our method for diagnosing the zygosity of infant twins demonstrated a high degree of accuracy.

Twin Analyses

Adjusted Activity Scores

Plomin and Foch (1980) suggest that it is important to adjust scores to eliminate the influence of covariance due to between-pair age differences on twin intraclass correlations. Thus, at both ages, age-adjusted activity scores were created for both CA and GA by regressing the composite actometer score on each age variable and using each twin's residuals as age-independent measures of AL. In a similar manner, measures of AL that were free from the effects of motor maturity were derived at each age by regressing the composite actometer score on the scaled motor score and using the residuals.

Preliminary Analyses

Prior to conducting statistical analyses to evaluate twin concordance, it is necessary to demonstrate that MZ and DZ twins do not represent two unique populations. To do this, for each variable at each age, one must evaluate whether the total means and variances are equal for each twin type. The finding of a significant difference between MZ and DZ means suggests an association between twin type and the variable being examined. Similarly, significant differences between variances provides evidence for an association between twin type and sources of variation (Christian, 1979). Consequently, the absence of differences between total means and variances is a necessary condition for the classical twin study (Plomin & Foch, 1980).

To test this critical assumption, a series of statistical analyses described by Christian (1979), were performed. For each variable of interest, t' tests based on the nested structure of twin data were conducted to evaluate for mean differences between twin types, and F' tests were conducted to test for homogeneity of variances between both twin types. Subsequent statistical tests of the genetic hypothesis were performed only for those variables that met the criteria of no mean differences between the MZ and DZ twin groups.

Intraclass Correlations

As indices of similarity between members of twin pairs, intraclass correlations were calculated for both twin types. This statistic estimates the proportion of total variance that is shared by twin siblings. For MZ twins, this proportion includes environmental variance and genetic variance, whereas for DZ twins, this proportion includes environmental variance but only one-half the additive genetic variance. Thus, given the assumption of equal environments for MZ and DZ twins, any difference between the intraclass correlations of MZ and DZ twins is considered to reflect a difference in genetic variance.

In the present study, MZ and DZ intraclass correlations were calculated for each variable of interest at both age levels. This calculation involves a one-way analysis of variance (ANOVA) for each twin type, with twin pair as the single factor (2 subjects per cell). Using the mean squares from the ANOVA, intraclass correlations for each twin type are then calculated as:

$$R = [\mathrm{MSB} - \mathrm{MSW}]/[\mathrm{MSB} + \mathrm{MSW}],$$

where MSB is the mean square (variance) between twin pairs, and MSW is the mean square within twin pairs.

Intraclass correlations estimate the degree of co-twin similarity for each twin type. Genetic influences are indicated when MZ co-twins are significantly more similar than DZ co-twins. Because there is some dispute over the best way to test for genetic variance, this was evaluated using two methods. First, the difference between MZ and DZ intraclass correlations was tested using Fisher's r to z transformation. Second, following Christian (1979), F_w ratios of DZ within-mean squares to MZ within-mean squares ($df = N_{DZ}, N_{MZ}$) were used to evaluate the question of significant genetic variance when the total variances of MZ and DZ twins did not significantly differ. As recommended, adjusted F_w' ratios were used to test genetic variance when the probability of unequal total variances exceeded .2 (Christian, Kang, & Norton, 1974; Christian, 1979). The adjusted F_w' is computed as:

$$F_w' = (MSW_{DZ} + MSB_{MZ})/(MSB_{DZ} + MSW_{MZ}),$$

with degrees of freedom as per Cochran (1951).

Occasionally, the two tests of genetic variance produced different outcomes. In these cases, a significant genetic effect was interpreted when the effect was significant at $p < .05$ for one method and at $p < .10$ for the other.

Genetic Analysis of Change and Continuity

Twin Concordance for Change

An analysis of twin concordance for change is possible for those variables that were measured in the same manner at both ages. For each child, change scores were calculated (Age 2 score minus Age 1 score). Because there was some variability in the length of interval between the initial and follow-up studies, that may contribute to between-pair variance, slope scores denoting rate of change were also calculated (change score divided by interval). Twin intraclass correlations for change[1] and slope scores were then derived via the analysis-of-variance method, and tests of genetic variance were conducted.

---

[1]This analysis is equivalent to Wilson's (1979b) repeated-measures ANOVA for longitudinal twin data for two measurement occasions.

<u>Cross-twin Intraclass Correlations</u>

To evaluate genetic influences on phenotypic continuity, cross-twin intraclass correlations were calculated for the variables in the change analyses. The cross-twin intraclass correlation is a cross-twin, cross-age intraclass correlation, where each twin's Age 1 score is cross correlated with their co-twin's Age 2 score. Because this involves two different variables (Age 1 and Age 2), the intraclass correlation cannot be computed through an analysis-of variance, and was calculated using the double-entry method, a technique that involves double entering the data so that each twin serves as both independent and dependent variables.[2] MZ cross-twin intraclass correlations that are significantly greater that the DZ cross-twin intraclass correlations provide evidence that heritability mediates phenotypic stability. This was tested using Fisher's *r* to *z* transformation.

[2]The double-entry and analysis of variance methods of calculating intraclass correlations yield similar results; however, the analysis of variance method is preferable because it provides the mean squares necessary for the F_w test of genetic variance.

RESULTS

Descriptive Statistics

Table 3 lists the means and standard deviations for the core variables across the two measurement occasions. Several features are worth noting. First, the mean of the scaled motor score during infancy, although within the average range, is below the BSID normalized mean of 100. This is consistent with Wilson and Harpring's (1972) finding of developmental lags in twin mental and motor development on the BSID, and likely reflects the effects of prematurity. This lag in motor development is not, however, apparent in early childhood. The mean scaled motor score on the MSCA suggests that the twin sample was quite representative of the normal population. The twin mean of 48.67 was within the standard error (±4) of the standardization mean of 49.9 for 3-year-old singletons (McCarthy, 1972).

An examination of those variables that were measured in the same manner on both occasions (actometer-assessed AL, weight, length/height, ponderal index, and head circumference), reveals that considerable change has taken place. Activity level shows a large increase of approximately 2 standard deviations. Physical change is even more prodigious. Increases for weight, length/height, and head circumference were in the magnitude of at least 4 standard deviations. A large *decrease* in the order of 4 standard deviations was found for ponderal index, a measure of relative body fat.

To evaluate the significance of these developmental changes, and to explore for possible sex differences, for each variable, a repeated measures ANOVA was

Table 3. Means and Standard Deviations for Core Variables at 7 and 36 Months of Age

Variable	Mean	SD
7 Months		
Composite actometer score	1.62	0.19
Parent rated AL (IBQ)	4.62	1.05
Scaled motor score (BSID)	92.23	11.56
Raw motor score (BSID)	30.04	5.84
Weight (kg)	7.96	1.20
Length (cm)	67.04	2.85
Ponderal index	2.63	0.26
Head circumference (cm)	44.28	1.70
36 Months		
Composite actometer score	2.07	0.12
Parent rated AL (TBAQ)	3.82	0.86
Scaled motor score (MSCA)	48.67	6.64
Raw motor score (MSCA)	13.17	3.84
Weight (kg)	13.98	1.73
Height (cm)	93.61	3.79
Ponderal index	1.70	0.16
Head circumference (cm)	50.03	1.44

Note. N = 76 - 80 individuals.

conducted with age as a within-subjects variable and gender as a between-subjects variable. To avoid problems arising from the paired nature of the data, this analysis was conducted twice -- once for twins A and once for twins B (see Table 4). Generally, the two analyses replicate each other. For all variables, large and highly significant, age effects are apparent for both twins A and B. Both twin groups

Table 4. Significant F Statistics for Age and Sex Effects by Twin

Variable	Twins A			Twins B		
	Age	Sex	Age x Sex	Age	Sex	Age x Sex
AL	181.26^{**}	--	--	195.56^{**}	7.57^{*}	--
Weight	1025.96^{**}	--	--	1024.86^{**}	--	--
Length/Height	2116.10^{**}	--	--	2101.76^{**}	--	--
Ponderal Index	652.38^{**}	--	--	676.24^{**}	--	--
Head Circumference	1315.75^{**}	11.3^{*}	--	1283.95^{**}	5.19^{*}	--

Note. Ns for Twins A and Twins B = 39 - 40.
$^{*}p < .05$.
$^{*}p < .0001$.

displayed a significant gender effect for head circumference, with males having larger heads than females. In addition, neither twin group displayed any significant interactions between age and gender. The one discrepancy between the MANOVA analyses for the two twin groups was the finding that although there was a significant sex effect for actometer-assessed AL for twins B, the sex effect did not approach significance for twins A.

To clarify this issue of sex differences in activity level, for both ages, analyses based on the nested structure of twin data were conducted, using the sex by family interaction as the error term in a one-way analysis of variance evaluating the sex effect. The sex difference in actometer-assessed AL was neither significant in infancy, $F(1, 38) = 3.32$, $p < .08$, nor in early childhood, $F(1, 38) = 0.92$, $p < .34$. The sex effects were, however, in the expected direction with males being more active than females at both ages ($M_{Males\ 1}$ = 1.67, $M_{Females\ 1}$ = 1.57, ES_1 = .53; $M_{Males\ 2}$ = 2.09, $M_{Females\ 2}$ = 2.06, ES_2 = .28).

Phenotypic Stability

Stability correlations for the full sample from 7 months to 36 months of age are presented in Table 5. Because our data consists of twin pairs, the significance levels of the correlations could be influenced by a lack of independence. Consequently, we used a reduced degrees of freedom based on the number of twin pairs ($df = N_{pairs} - 2$) to evaluate significance. This conservative procedure avoids inflating the probabilities of finding significant associations.[3]

The stability of activity level varies according to measurement procedure. When assessed mechanically, the AL correlation is low and nonsignificant, suggesting little continuity in individual differences from infancy to early childhood. However, the higher, statistically significant, cross-age correlation for parent rated activity level suggests moderate phenotypic continuity. As predicted, the infant measure of motor development is a relatively poor predictor of motor skills in early childhood. Cross-age correlations for both, raw motor scores and scaled motor scores, were weak and failed to reach significance in our twin sample. Despite the considerable mean level changes apparent in the anthropometric measures of weight, length/height, ponderal index, and head circumference, the moderate to high correlations from infancy to early

---

[3]Performing the correlation analyses separately on Twins A and Twins B, yielded patterns of correlations that paralleled each other and that of the full sample, in terms of both magnitude and significance level.

Table 5. Phenotypic Stability Correlations from 7 to 36 Months of Age

Variable	Cross-age *r*
Actometer-assessed AL	.15
Parent-rated AL	.41*
Raw motor score	.15
Scaled motor score	.23
Weight	.72**
Length/height	.43*
Ponderal index	.52*
Head circumference	.81**

Note. *df* = 36 - 38.
*$p < .01$.
**$p < .0001$.

childhood indicate that there is substantial phenotypic stability. Thus, amongst the flux of physical development, there is much constancy in terms of one's ordinal position within the group.

Genetic Influences at 7 and 36 Months

Twin intraclass correlations and tests of genetic variance are presented in Table 6. One variable, the 36-month scaled motor score, did not meet the assumptions for the genetic analysis because preliminary analyses revealed a significant mean level difference between MZ and DZ twin groups; the DZ twins scored significantly higher than the MZ twins. However, there was no significant mean difference between raw motor scores of MZ and DZ twins. This pattern is not unusual, when one considers that MZ twins are chronologically but not gestationally older. Although the intraclass

Table 6. Intraclass Correlations, Standard Errors, and Statistical Tests of Genetic Variance at 7 and 36 Months of Age

	7 Months					36 Months				
	R_{MZ}[a]		R_{DZ}[b]		F_w	R_{MZ}[a]		R_{DZ}[b]		F_w[c]
Actometer assessed AL										
Composite actometer score	**.88****	±.04	**.43**	±.24	4.08**	**.76***	±.08	**.38**	±.26	2.98**
CA-adjusted AL	**.85****	±.05	**.37**	±.26	4.08**	**.77***	±.08	**.25**	±.28	2.98**
GA-adjusted AL	**.86****	±.05	**.32**	±.27	4.08**	**.77***	±.08	**.31**	±.27	2.98**
Motor-adjusted AL	**.89****	±.04	**.40**	±.25	4.46**	**.76**	±.09	**.38**	±.26	2.98**
Parent rated AL	**.72****	±.09	**-.02**	±.30	3.43**	**.93*****	±.03	**.23**	±.29	14.60***
Motor development										
Raw motor score	**.98**	±.01	**.96*****	±.03	1.37	**.78***	±.08	**.33**	±.27	2.30*
Scaled motor score	**.92**	±.03	**.86*****	±.08	1.67	**.79***	±.08	**.34**	±.27	--
Anthropometric measures										
Weight	**.90***	±.04	**.69****	±.16	1.92	**.93*****	±.03	**.48***	±.23	5.57***
Length/height	**.90***	±.04	**.69****	±.16	2.37*	**.94****	±.02	**.54***	±.21	6.13***
Ponderal index	**.71**	±.10	**.30**	±.27	2.03	**.80***	±.07	**.38**	±.26	4.26***
Head circumference	**.89**	±.04	**.71****	±.15	2.44*d	**.91****	±.03	**.41**	±.25	4.12***

Note. [a]Asterisk in R_{MZ} column indicates that R_{MZ} is significantly greater than R_{DZ}.
[b]Asterisk in R_{DZ} column indicates that R_{DZ} is significantly greater than zero.
[c]Missing value in F_w column indicates that the genetic hypothesis was not tested.
[d]Adjusted F_w' was used to test genetic variance.
* $p < .05$. ** $p < .01$. *** $p < .001$.

correlations for the 36-month scaled motor score are presented in Table 6, tests of genetic influence were not conducted because of the mean level difference. The assumption of homogeneity of variance between twin groups was met for all variables except 7-month head circumference. Consequently, when testing for significant genetic variance on this variable, an adjusted F_w' was employed.

At both 7 and 36 months of age, the MZ intraclass correlations are high and significantly different from zero. In contrast, the DZ correlations are generally lower, with only the correlations for 7-month raw motor score, 7-month scaled motor score, 7- and 36-month weight scores, and 7- and 36-month length scores being significantly different from zero. The Fisher's z test of significant differences between MZ and DZ intraclass correlations and the F_w ratios agreed with each other for all variables with three exceptions. For the 7-month weight variable, the z test was significant, but the F_w ratio only approached significance ($p < .08$). Similarly, the appropriate F_w ratios for head circumference and the 36-month measure of motor-adjusted AL were significant, whereas the z tests approached significance ($p < .08$, $p < .06$, respectively). These differences are likely a consequence of the small sample size and, possibly, some undetected heterogeneity of variance.

Actometer-Assessed AL

Composite Actometer Score

At both 7 and 36 months of age, the intraclass correlations for the composite actometer scores provide evidence of genetic influences on motor activity level. On this objective measure, MZ co-twins were significantly more similar in AL than DZ co-twins in infancy and early childhood. Moreover, the twin concordances conformed to the classic twin model, with the DZ correlations being approximately one-half the MZ correlations.

Adjusting for Age

Chronological (CA) and gestationally-adjusted (GA) age were significantly correlated with actometer-assessed AL in infancy, $r(38) = .40$, $p < .01$; and $r(38) = .41$, $p < .01$, respectively, but not in early childhood $r(38) = -.26$, $p > .1$; and $r(38) = -.15$, $p > .3$, respectively. Removal of age effects from the composite actometer scores resulted in little change in the overall pattern of MZ and DZ co-twin resemblances. The intraclass correlations for CA-adjusted and GA-adjusted activity level continue to show evidence of a significant genetic influence at both ages. Thus, for the age ranges studied, age does not appear to significantly mediate the degree of co-twin similarity.

Adjusting for Motor Maturity

During infancy, the composite actometer score was significantly correlated with the raw motor score on the Bayley Motor Scale, $r(38) = .49$, $p < .01$, but not with the scaled motor score (PDI), $r(38) = .21$, $p > .18$. This pattern suggests that age is mediating the significant correlation between activity level and the raw motor score on the Bayley. Therefore, when adjusting AL for the effects of motor maturity, we regressed composite actometer score on the scaled motor score to obtain a motor-adjusted measure of AL that was not confounded with age effects. Adjusting the actometer score to eliminate the influences of motor maturity did not substantially alter the pattern of intraclass correlations.

In early childhood, neither the raw nor scaled motor scores were significantly related to activity level, $r(36) = -.07$, $p > .6$, $r(36) = -.00$, $p > .9$, respectively. Given these trivial, nonsignificant correlations, it was not surprising to find that at 36 months, there was no change in the MZ and DZ intraclass correlations when activity level was adjusted for motor maturity.

Parent-Rated AL

Like the composite actometer scores, parent ratings of activity on the Infant Behavior Questionnaire (IBQ) and on the Toddler Behavior Assessment Questionnaire (TBAQ) yield evidence of genetic influences on AL. At both ages, the pattern of high MZ correlations and low DZ correlations (negative in the case of the infant measure) violates the classic twin model and suggests the possible presence of rater bias. However, with our small sample the DZ correlations have large standard errors, and when these are taken into consideration, the DZ correlations do not differ significantly from half the MZ correlations, as would be necessary to clearly show a rater bias.

Parent ratings of activity level were positively related to the composite actometer score in infancy, $r(38) = .36$, $p < .05$; but not in early childhood, $r(38) = .07$, $p > .6$. Thus, moderate convergent validity was demonstrated for the Activity scale of Rothbart's (1981) IBQ, but not for Goldsmith's (1987) TBAQ.

Motor Development

Infant motor development shows no evidence of a genetic influence in our sample. For both the BSID raw motor score and the BSID scaled motor score, there was no significant difference between the very high intraclass correlations demonstrated by both MZ and DZ twins. Furthermore, the finding that the DZ co-twin resemblance is substantially higher than would be predicted by the genetic hypothesis suggests that environmental influences are enhancing co-twin similarity for this measure.

Motor development displays a different pattern of co-twin resemblance during early childhood. The raw motor score on the *McCarthy Scales of Children's Abilities* evinces evidence of a genetic influence. For this general measure of motor competence, MZ co-twins are significantly more similar than DZ co-twins, and twin intraclass correlations conform to the classic twin model.

Anthropometric Measures

Overall, across both ages, our sample conforms to genetic expectations for anthropometric indices. Significant genetic variance was indicated for the infant measures of weight, length, and head circumference. For these variables, the MZ twins displayed a high degree of co-twin resemblance in contrast to the more moderate DZ co-twin resemblance. For infant ponderal index, both tests of genetic influence approached significance ($p < .06$) suggesting that genetic influences may also be operating for this variable.

At follow-up, all anthropometric measures yield evidence of significant genetic influence. A pattern of high MZ intraclass correlations and DZ correlations that are approximately half as large, was evident for weight, height, ponderal index, and head circumference.

Differential Heritability

Heritability estimates are not presented because of their instability with small samples, thus precluding statistical tests of differential heritability. Despite this, some tentative comments can be made regarding the patterns of twin correlations from 7 to 36 months and their implications for changes in heritability. For most measures, the differences between the MZ and DZ correlations are similar across age, providing little support for differential heritability. In contrast, the motor development raw score and head circumference show a large increase in the difference between MZ and DZ correlations from infancy to early childhood. The DZ twins in particular, would appear to be becoming more dissimilar with age. This pattern implies increasing heritability, however; because the standard errors of the DZ correlations tend to be large, these results are merely suggestive.

Genetic Influences on Change and Continuity

Analysis of Change

Change scores and slope scores were calculated for actometer-assessed AL, weight, length/height, ponderal index, and head circumference. Preliminary analyses revealed significant variance heterogeneity for change in weight, length and ponderal index; and slope of change in ponderal index. In addition, there were significant twin mean differences for change in weight, and for slope of change in weight and head circumference. For these two variables, DZ twins displayed a greater amount of developmental change.

Twin intraclass correlations for change scores and slope scores are presented in Table 7. Overall, the MZ correlations for change and slope are similar in magnitude and significance, whereas the DZ correlations show some divergence. The DZ change correlations are moderate to low, with change in length/height and head circumference being significantly greater than zero; and change in change in AL approaching significance ($p < .06$). By comparison, ponderal index was the only DZ slope correlation that did not reach significance.

For both change and slope, the Fisher's z test and the F_w/F_w' ratios concurred in detecting significant genetic variance. However, the overall pattern of significant differences between MZ and DZ co-twin similarity differs across change and slope methods of assessing change. Genetic influences were suggested for relative change in AL, length/height, ponderal index, and head circumference. In contrast, only AL and ponderal index demonstrated significant genetic variance for rate of change or slope scores. These differences in genetic outcomes suggests that because change scores fail to take into account variability in the length of interval, the between-pair variance (MSB) is reduced, resulting in decreased DZ co-twin similarity. The resemblance of MZ co-twins on relative change scores appears to be less affected by variability in

Table 7. **Intraclass Correlations, Standard Errors, and Statistical Tests of Genetic Variance for Change and Slope Scores from 7 to 36 Months of Age**

	Change Scores					Slope Scores				
	R_{MZ}[a]		R_{DZ}[b]		F_w[c]	R_{MZ}		R_{DZ}		F_w[c]
Actometer assessed AL	**.85***	±.05	**.44**	±.24	3.35**	**.84***	±.06	**.48***	±.23	3.43**
Weight	**.92*****	±.03	**.19**	±.29	--	**.91****	±.03	**.54***	±.21	--
Length/height	**.93****	±.03	**.56***	±.21	4.25^{**d}	**.91**	±.04	**.73***	±.14	1.58
Ponderal index	**.71***	±.10	**.10**	±.30	2.80^{**d}	**.68***	±.10	**.02**	±.30	2.82^{*d}
Head circumference	**.83***	±.06	**.49***	±.23	2.14*	**.78**	±.08	**.51***	±.22	--

Note. [a]Asterisk in R_{MZ} columns indicates that R_{MZ} is significantly greater than R_{DZ}.
[b]Asterisk in the R_{DZ} column indicates that R_{DZ} is significantly greater than zero.
[c]Missing value in F_w column indicates that the genetic hypothesis was not tested.
[d]Adjusted F_w' was used to test genetic variance.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

interval length, presumably as a result of the larger MZ sample. Therefore, in the present study, interval length is an important source of nongenetic source of variance that when not accounted for, reduces the resemblance of the DZ twins, and hence, serves to inflate the estimates of genetic variance. Consequently, the significant genetic influences indicated for relative change should be viewed skeptically.

Because they control for interval length, slope scores are more appropriate for evaluating genetic influences on developmental change. The finding that MZ co-twins are significantly more similar than DZ co-twins for rate of change in actometer-assessed AL and ponderal index suggests that, for these variables, developmental change is regulated by genetic factors.

Analysis of Continuity

Cross-twin intraclass correlations are presented in Table 8. Genetic influences on phenotypic continuity are suggested when the MZ cross-twin correlation is significantly greater than the DZ cross-twin correlation. This was not apparent for any variable. However, with low to moderate phenotypic correlations, low cross-twin correlations, and a small sample, a continuity analysis has little statistical power to detect significant differences between MZ and DZ twin groups. Despite this, some speculative comments regarding genetic continuity can be made on the basis of the overall pattern of results.

The cross-twin correlations for activity level, weight, and head circumference are consistent with the twin model of additive genetic effects. That is, the MZ

Table 8. Cross-twin Correlations and Standard Errors for Genetic Continuity From 7 to 36 Months of Age

	Cross-twin Correlations			
	R_{MZ}[a]		R_{DZ}[b]	
Actometer assessed AL	**.13**†	±.19	**.05**	±.30
Weight	**.41**	±.16	**.17**	±.29
Length/height	**.28**†	±.18	**.48***	±.23
Ponderal index	**.32**	±.18	**.29**	±.28
Head circumference	**.76**	±.08	**.41**	±.25

Note. [a]Dagger in R_{MZ} column indicates that R_{MZ} is not significantly different from zero. [b]Asterisk in the R_{DZ} column indicates that R_{DZ} is significantly greater than zero.
* $p < .05$.

correlations are approximately twice those of the DZ group. In addition, the difference between the MZ and DZ cross-twin correlations for head circumference, the variable with the highest degree of stability, approached significance ($p < .07$). Thus, the general pattern of MZ and DZ cross-twin correlations hint that genetic influences may be contributing to the phenotypic continuity of these variables.

In contrast, cross-twin correlations for ponderal index and length/height imply that environmental factors may be substantially mediating the phenotypic continuity of these anthropometric variables. For example, there was little difference between the MZ and DZ cross-twin correlations for ponderal index. Thus, it would appear that the number of genes in common is not related to the magnitude of the correlation, as would be required to demonstrate a genetic influence. However, given the large standard errors associated with the correlations, and the power limitations outlined above, the results are biased towards proving the null hypothesis; hence this conclusion is merely tentative.

An unexpected outcome was the finding that for length/height, the MZ cross-twin correlation was nonsignificant, whereas the DZ cross-twin correlation was significantly different from zero. This result makes sense when one looks at the phenotypic stability of MZ and DZ twins separately. Length/height is the only variable for which the MZ and DZ stabilities significantly differ ($r_{MZ} = .30$, $r_{DZ} = .83$, $p_{difference} < .05$). Thus, MZ twins, as a group, experience a greater reordering of individual differences across age. Because the MZ stability is low, the MZ cross-twin correlation will also be low. Note, however, that the magnitude of the MZ stability of length/height is very close to the MZ cross-twin correlation. Thus, for MZ twins, one could predict with equal accuracy, a twin's time 2 height from their own or from their co-twin's length at infancy.

DISCUSSION

Developmental behavior genetics research enables one to go beyond the simple description and prediction of individual differences in temperament to explore the *etiology* of individual differences and of developmental changes in these individual differences. However, few longitudinal studies have explored continuity and change in temperament from a developmental behavior genetic perspective. Those that have, have relied on parent or observer ratings of behavior, which have limitations as discussed earlier. The present study extends previous research through the use of a mechanical measure of activity level to evaluate the importance of genetic influences on the continuity and change of individual differences in activity level from 7 to 36 months of age. The study's two major hypotheses were supported: Objectively assessed individual differences in activity level evince genetic influences in infancy *and* early childhood, and developmental *change* in activity level appears to be genetically mediated.

As predicted, at both 7 and 36 months of age, identical twins were significantly more similar than fraternal twins on an in-home, 48-hour mechanical measure of motor activity. Moreover, with this measure, MZ and DZ concordances conformed to the classic twin model of additive genetic effects at both ages. Thus, although at 36 months of age, the young child is encountering and interacting with more diverse environments, this increase in environmental variance does not negate the importance of genetic inputs to motor activity level. Indeed, previous temperament research suggests that when developmental changes in genetic influences are indicated, it is in the direction of *increased* genetic variance (e.g., Braungart, et al., 1992; Buss et al., 1973; Stevenson & Fielding, 1985; Torgersen, 1981; Torgersen & Kringlen, 1978). In the present study, the difference between the MZ and DZ intraclass correlations for actometer-assessed AL was similar across age, suggesting little differential heritability across the intervening 2½ years. However, given the large standard errors surrounding the correlations, conclusions regarding differential heritability must be made cautiously. The pattern of results is, however, consistent with a recent developmental meta-analysis of twin studies by McCartney, Harris, and Bernieri (1990) that found MZ and DZ intraclass correlations for AL decreased to a similar extent with age, and that there was no significant relation between age and heritability estimates for AL.

To the developmentalist, a focus on genetic influences on developmental change and continuity is more interesting than the presence of genetic influences at a single age, or the findings of differential heritability across age, because the change/continuity data speak to the question of how developmental change takes place. From 7 to 36 months of age, actometer-assessed activity level demonstrated considerable developmental change. Over this interval, the sample displayed a significant increase in mean level of AL, and a re-ordering of individual differences. The genetic analysis of change scores revealed that MZ concordance for change was significantly greater than that of DZ twins. Thus, the observed developmental change in AL would appear to be partially regulated by genetic influences.

Evidence for genetic patterning of change in AL has been demonstrated in two previous twin studies of infant temperament. In the LTS, change in observer-rated IBR activity from 12 to 24 months was significantly heritable (Matheny, 1983). Similarly, in the MALTS, parent ratings of AL from 14 to 20 months displayed a genetic influence (Plomin et al., 1992). The results of the present study support these findings while making several unique contributions to the existing literature. First, the present research spans two developmental periods, specifically, infancy and childhood. Although, the notion of a transition from infancy to early childhood is generally accepted, previous temperament research has tended to focus on change *within* a single developmental period (Plomin, DeFries, & Fulker, 1988). The assessment of activity level at 7 and 36 months of age enables an analysis of the considerable change characterized by the transition from infancy to early childhood. The finding of a genetic influence on change in AL across this interval begins to address the etiology of developmental change.

Second, prior studies examining genetic influences on developmental change in temperament have relied on parent and observer ratings. Parent ratings of temperament have been shown to be prone to rater biases (e.g., Neale & Stevenson, 1989); whereas observer ratings, although more objective, permit only limited behavioral sampling. In addition, both measures may exhibit differential reliability and validity across age. The use of motion recorders to assess activity level circumvents the problems associated with parent and observer ratings. This measurement procedure demonstrates ecological validity and cross situational generality, ensures measurement equivalency across age, and is clearly an objective measure of motor activity. Thus, the present finding of genetic influences on developmental change in actometer-assessed activity level substantially strengthens previous findings.

A related issue has to do with the measurement of change in activity level. According to contemporary temperament theories, temperament refers to early appearing individual differences in behavioral tendencies that have a constitutional basis, and that demonstrate continuity in expression across time and situations (Goldsmith et al., 1987). Thus, by definition, activity level is conceptualized as a stable dimension. Consequently, measures of activity that are rooted in temperament theory will be biased toward stability -- this includes most rating scales. Indeed, stability estimates are included as an important psychometric property when evaluating the reliability and validity of temperament instruments (e.g., Bates, 1986; Hubert, Wachs, Peters-Martin, Gandour, 1982). Therefore, traditional temperament measures may be insensitive, if not inadequate, to detect ontological change.

Moreover, with rating measures, parents and observers do not *directly* evaluate changes in temperament; they simply rate the temperament of children on repeated occasions (Loehlin, Horn, & Willerman, 1990). Change is then inferred by a difference in ratings across age. However, considerable developmental change can go undetected with this approach. For example, one IBR item asks the observer to rate the infant's amount of gross bodily movements observed during administration of the BSID. Although the activity level of a moderately active 1-year-old most certainly differs from that of a moderately active 2-year-old, both would receive a score of 5 on this item because their behavior is judged relative to their same-aged peers. Thus, an instrument designed to focus on stability is unlikely to provide a good measure of developmental change. This is not a problem with the actometer because it is an absolute measurement referencing system (Aftanas, 1986). That is, it provides a count measure of gross motor movement over a given interval, in this case 48-hours. The difference in the movement counts at 7 and 36 months of age, therefore, yields an absolute measure of change in AL across age. For this reason, the actometer may be more appropriate than rating scales for assessing developmental change in activity level.

Because the actometer is more sensitive to detecting developmental change, it may also provide a more rigorous test of genetic contributions to phenotypic stability. Mechanically-assessed AL demonstrated significant genetic variance at both 7 and 36 months of age, yet phenotypic stability was low. Given little phenotypic stability, the finding of no significant genetic influences on the phenotypic stability of AL was not surprising. Low phenotypic stability suggests that the genetic and environmental influences that govern a characteristic at one age do not correlate across the two ages (Plomin, 1986a). This leads to the inference that the genes that operate on individual differences in AL at 7 months of age differ from those that operate at 36 months of age, and therefore, implies that genetic change has taken place. Nonetheless, there may be some undetected genetic contribution to phenotypic continuity. When phenotypic stability is low, as was the case in the present study, MZ and DZ cross-twin correlations are also low, and consequently, very large samples are required to detect significant differences between twin types. The pattern of an MZ cross-twin correlation, .13, that was twice as large as that of the DZ twin group, .05, is consistent with the genetic hypothesis, and may indicate that genetic influences are moderating whatever phenotypic stability exists. Indeed, this would be congruous with Plomin et al.'s (1992) finding that genetic factors account for nearly all the phenotypic stability in IBR-rated activity across 14 to 20 months of age.

It should be acknowledged, however, that the present study spans a period of time in which there are momentous changes in locomotor behavior. For example, between 7 and 36 months of age, the child goes from pre-crawling behaviors to co-ordinated walking. In light of these vast changes, low age-to-age stability in motor activity does not seem unusual. It is likely the phenotypic stability of AL will be higher when the child is not crossing stage boundaries. An analysis of genetic contributions to continuity might prove more fruitful within an intra-stage context.

Because parent-rated AL was assessed with different questionnaires at 7 and 36 months of age, a full analysis of continuity and change is not possible. Nevertheless, the parent ratings of activity produced some interesting results. In contrast to actometer-assessed AL, parent ratings of activity demonstrated significant stability across the two different temperament questionnaires. This pattern of results is consistent with previous research which finds that, as compared to observer ratings of AL, parent ratings yield higher stability correlations (e.g., Plomin et al., 1988; Plomin et al., 1992; Wilson & Matheny, 1986). In the present study, parent ratings of AL demonstrated convergent validity with the actometer measure during infancy, but not in early childhood. That is, parent ratings of activity on the TBAQ were not related to mechanically assessed activity. Taken together, these results suggest that the apparent stability in activity level is in the eye of the rater rather than in the behavior of the child.

Possible rater bias was also suggested from the analysis of genetic influence on parent-rated AL. Although, parent ratings of activity demonstrated evidence of genetic influences at 7 and 36 months of age, the problem of "too low" DZ correlations was apparent for both the IBQ and TBAQ measures. If a contrast bias is operative, the DZ correlation should be less than half the MZ correlation. This pattern emerges in the present study; however, given the large standard errors surrounding the DZ correlations, they do not *significantly* differ from half the MZ correlations. Consequently, a definitive statement about rater bias cannot be made without replication with a substantially larger sample.

Temperament research has generally ignored the possibility that differences in developmental rate may be responsible for temperamental differences, a possibility that seems particularly plausible in the case of activity level (Eaton, in press). Therefore, measures of motor development were included in the present study as indices of motor maturity, and provided a way of evaluating whether concordance in motor maturity could account for concordance in activity level. During infancy, there was no evidence of genetic influences on the BSID motor scale; both twin types displayed very high within-pair similarity. This suggests that shared environmental factors are operating to enhance the similarity of co-twins. An alternate explanation proposing that age might be mediating the observed resemblance between co-twins, was considered; however, adjusting the raw and scaled PDI scores for the effects of chronological and gestationally-adjusted age did not change the pattern of co-twin similarity. Although not as extreme, Wilson and Harpring (1972) also report high MZ and DZ co-twin resemblance on the PDI motor scale in a sample of 6-month-olds. In addition, they found evidence of significant genetic variance. Moreover, sibling data from the Colorado Adoption Project suggest greater genetic influence on the PDI than do Wilson and Harpring's twin data (Plomin et al., 1988). Thus, the failure to find significant genetic influence in the present study may be a result of low statistical power.

Motor development in early childhood did, however, display genetic influences. At 36 months of age, the DZ intraclass correlation was significantly lower than that of the MZ twin group. In addition, co-twin resemblances conformed more closely to the classic twin model. Although co-twin similarity in motor development appeared to decrease across age for both twin groups, the DZ co-twin resemblance appeared to decrease to a greater extent. This pattern is suggestive of increasing heritability, and it may be that with development, genetically influenced individual differences in motor ability are unmasked. Although this is a reasonable conclusion, because different scales were used to assess motor development at each age, the apparent increase in heritability might well arise from measurement artifact. That is, the infancy measure of motor maturity may be less sensitive in detecting intra-pair differences, resulting in the very high infant DZ correlations and the consequent attenuation of genetic effects.

Do the observed co-twin similarities in motor development have any bearing on similarity in activity level? The answer would appear to be no. In infancy PDI score was positively associated with AL, yet adjusting the actometer score for the effects of motor maturity, did not alter the pattern of co-twin resemblance. At the 36-month assessment, moreover, the predicted relation between motor development and AL was not found, so adjusting the 36-month actometer scores for motor effects produced no change in MZ and DZ intraclass correlations. Thus, it would appear that similarity in activity level is not simply an artifact of similarity in motor maturity, and consequently, activity level represents a maturation-independent individual differences dimension.

Anthropometric indices were included in the present study to check that the sample conformed to developmental and genetic expectations. As expected, there was much mean level change in weight, length/height, ponderal index and head circumference, yet there was also significant stability from 7 to 36 months of age.

Plomin et al., (1988) found height and weight to be remarkably stable in infancy and early childhood, and the present results are in agreement. At both ages, MZ twins were more similar than DZ twins for all measured physical characteristics. Although the present results are consistent with previous research suggesting genetic influences on physical development across infancy and early childhood (e.g., Plomin et al., 1988; Wilson, 1979a, 1979b), the clear trend toward increasing heritability of height and weight found in previous research was not apparent. Twin correlations for head circumference were, however, suggestive of differential heritability.

The analysis of change on the anthropometric indices revealed significant MZ and DZ group differences. The DZ twin group displayed a greater mean rate of change, as indicated by slope scores, in weight and head circumference. The variance of the slope scores for ponderal index was significantly greater for the MZ than DZ twin group. In addition, the DZ twins demonstrated significantly higher stability for length/height. It is noteworthy that, of all the variables assessed, it is only these physical growth variables that displayed significant zygosity group mean differences. Such differences imply violations of the equal environments assumption, and prenatal influences are a likely source of such environmental inequality. Approximately 70% of MZ twin pairs share a monochorionic placenta, and such placentas tend to be prone to unequal nutrition of twins (Plomin, DeFries, & McClearn, 1990; Wilson, 1979b). Wilson (1979b) found that although at birth, MZ twins were significantly *less* concordant for length than DZ twins, their similarity increased until the preschool-age.

Thus, in the present study, group differences could reflect the residual effects of perinatal factors. For example, catch-up growth in MZ twins would explain zygosity group differences in stability, and differences in rate of growth might denote enduring prenatal influences and limitations in recuperative power, especially for MZ twin pairs with large differences in birth weight.

These apparent growth differences between MZ and DZ twin groups made the genetic analysis of change and continuity less clear for the anthropometric measures. Overall, MZ co-twins are more concordant than DZ co-twins for physical change. Wilson (1979a) reports a similar pattern of results for his analysis of twin trend correlations for longitudinal profiles of height and weight, and suggests that these results reveal "powerful chronogenetic influences on growth" (p. 104). Height and weight data from the Colorado Adoption Project suggest that the phenotypic stability of height is almost entirely mediated by genetic factors, whereas for weight, both genetic and environmental factors affect stability. The present study does not allow a powerful test of genetic influences on phenotypic continuity, but the overall pattern of results hint that there may be some genetic overlap across age for weight and head circumference.

The limitations of this study should be noted. First, the classic twin study rests on the assumption of equal environmental variance for both twin types. Although previous research has demonstrated that this is a tenable proposition (e.g., Plomin, Willerman, & Loehlin, 1976; Scarr, 1966; Scarr & Carter-Saltzman, 1979; Torgersen

& Kringlen, 1978), the extent to which this assumption is violated would increase the likelihood of finding a significant difference between MZ and DZ co-twin similarity. Second, because zygosity was determined, in part, through physical similarity criteria, errors in classification may have occurred. Cross-validation of diagnoses in the present study suggested reasonable validity both within and across age. However, misdiagnoses of zygosity would act *against* the genetic hypothesis by reducing the differences between MZ and DZ resemblances. Third, because the sample in this study is small, the standard errors for intraclass correlations tend to be large; hence it is only possible to determine the *presence* of a genetic influence and not its magnitude. With small samples, twin studies can only detect large genetic effect; thus, a null finding is not evidence that genetic influences do not operate. Moreover, this study cannot provide a powerful statistical test of differential heritability or of genetic continuity. Tentative inferences regarding differential heritability and genetic continuity have been made on the basis of the overall patterns of MZ and DZ correlations, and should be viewed merely as interesting starting points for future research.

Despite these limitations, the present research has much to contribute to the study of temperament. The results of this first longitudinal twin study of objectively-assessed AL provide novel evidence of genetic influences on infant and child motor activity, and on developmental change in motor activity, and thus, begin to address the mechanisms involved in the emergence of individual differences. Clearly, genes play an important role in behavioral development. The question of how genes act to bring about developmental change in AL remains an issue for future research. Genetically mediated change in AL could indicate the presence of specific genes that are direct contributors to ontological change, or that the genes responsible for individual differences in infant AL differ from those that affect childhood AL. Because AL demonstrated significant genetic influences during infancy and early childhood, but little age-to-age stability, it is probable that the latter is true. Thus, ontological change in AL might result from genes switching on and off during development.

A more definitive answer about how genes operate on developmental change in temperament might one day be achieved through the merging of molecular and behavioral genetics. Recombinant DNA techniques now make it feasible to *directly* study DNA variation among individuals (Plomin, DeFries, & McClearn, 1990). With the identification of more DNA markers and the development of new research strategies, it may become possible to relate DNA variation to behavioral variation (Plomin, 1990). In the meantime, by indicating the presence of genetic influences on ontological change, the present study takes an important first step towards our understanding of the *process* of behavioral development, and the message it heralds is clear: Genes can, and should, be viewed as potent sources of developmental change.

REFERENCES

Aftanas, M. (1986). *Basic Principles in the Assessment and Measurement of Behavior.* Unpublished manuscript, University of Manitoba, Department of Psychology, Winnipeg, Manitoba.

Aftanas, M. (1988). Theories, models, and standard systems of measurement. *Applied Psychological Measurement, 12,* 325-338.

Bates, J. E. (1986). The measurement of temperament. In R. Plomin & J. Dunn (Eds.), *The study of temperament: Changes, continuities and challenges.* (pp. 1-12). Hillsdale, NJ: Erlbaum.

Bayley, N. (1969). *Manual for the Bayley Scales of Infant Development.* New York: The Psychological Corporation.

Braungart, J. M., Plomin, R., Defries, J. C., & Fulker, D. W. (1992). Genetic influence on tester-rated infant temperament as assessed by Bayley's Infant Behavior Record: Nonadoptive and adoptive siblings and twins. *Developmental Psychology, 28,* 40-47.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development.* New York: Wiley.

Buss, A. H., & Plomin, R. (1984). *Temperament: Early developing personality traits.* Hillsdale, NJ: Erlbaum.

Buss, A. H., Plomin, R., & Willerman, L. (1973). The inheritance of temperaments. *Journal of Personality, 41,* 513-524.

Carter-Saltzman, L., & Scarr, S. (1977). MZ or DZ? Only your blood grouping laboratory knows for sure. *Behavior Genetics, 7,* 273-280.

Christian, J. C. (1979). Testing twin means and estimating genetic variance. Basic methodology for the analysis of quantitative twin data. *Acta Geneticae Medicae et Gemellologiae, 28*, 35-40.

Christian, J. C., Kang, K. W., & Norton, J. A., Jr. (1974). Choice of an estimate of genetic variance from twin data. *American Journal of Human Genetics, 26,* 154-161.

Clarke, A. D. B., & Clarke, A. M. (1984). Constancy and change in the growth of human characteristics. *Journal of Child Psychology and Psychiatry, 25,* 191-210.

Cochran, W. G. (1951). Testing a relation among variances. *Biometrics, 7,* 17-32.

Cohen, D. J., Dibble, E., & Grawe, J. M. (1977). Fathers' and mothers' perceptions of children's personality. *Archives of General Psychiatry, 34*, 480-487.

Cohen, D. J., Dibble, E., Grawe, J. M., & Pollin, W. (1973). Separating identical from fraternal twins. *Archives of General Psychiatry, 29,* 465-469.

Cohen, D. J., Dibble, E., Grawe, J. M., & Pollin, W. (1975). Reliably separating identical from fraternal twins. *Archives of General Psychiatry, 32,* 1371-1375.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements.* New York: Wiley.

Cyphers, L. H., Phillips, K., Fulker, D. W., & Mrazek, D. A. (1990). Twin temperament during the transition from infancy to early childhood. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 392-397.

Dworkin, R. H., Burke, B. W., Maher, B. A., & Gottesman, I. I. (1976). A longitudinal study of the genetics of personality. *Journal of Personality and Social Psychology, 34,* 510-518.

Eaton, W. O. (1983). Measuring activity level with actometers: Reliability, validity, and arm length. *Child Development, 54,* 720-726.

Eaton, W. O. (in press). Temperament, development, and the five factor model: Lessons from activity level. In C. F. Halverson, G. A. Kohnstamm, R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood.* Hillsdale, N. J.: Erlbaum.

Eaton, W. O., & Dureski, C. M. (1986). Parent and actometer measures of motor activity level in the young infant. *Infant Behavior and Development, 9,* 383-393.

Eaton, W. O., McKeen, N. A., & Lam, C. (1988). Instrumented motor activity measurement of the young infant in the home: Validity and Reliability. *Infant Behavior and Development, 11,* 375-378.

Emde, R. N., Plomin, R., Robinson, J., Reznick, J. S., Campos, J., Corley, R., DeFries, J. C., Fulker, D. W., Kagan, J., & Zahn-Waxler, C. (in press). Temperament, emotion, and cognition at 14 months: The MacArthur Longitudinal Twin Study. *Child Development.*

Escalona, S. K. (1968). *The Roots of Individuality.* Chicago: Aldine.

Fagen, J .W., Singer, J. M., Ohr, P. S., Fleckenstein, L. K. (1987). Infant temperament and performance on the Bayley Scales of Infant Development at 4, 8, and 12 months of age. *Infant Behavior and Development, 10*, 505-512.

Falconer, D. S. (1981). *Introduction to Quantitative Genetics (2nd ed.).* New York: Longman Inc.

Fish, M., & Crockenberg, S. (1981). Correlates and antecedents of nine-month-old infant behavior and mother-infant interaction. *Infant Behavior and Development, 4,* 69-81.

Goldsmith, H. H. (1983). Genetic influences on personality from infancy to adulthood. *Child Development, 54,* 331-355.

Goldsmith, H. H. (1987). *The Toddler Temperament Behavior Assessment Questionnaire. A Preliminary Manual.* Unpublished manuscript, University of Oregon.

Goldsmith, H. H., & Campos, J. J. (1986). Fundamental issues in the study of early temperament: The Denver Twin Temperament Study. In M. E. Lamb, A. L. Brown, & B. Rogoff (Eds.), *Advances in developmental psychology* (Vol 4, pp. 231-283). Hillsdale, NJ: Erlbaum.

Goldsmith, H. H., & Gottesman, I. I. (1981). Origins of variation in behavioral style: A longitudinal study of temperament in young twins. *Child Development, 52,* 91-103.

Goldsmith, H. H., Rieser-Danner, L. A., & Briggs, S. (1991). Evaluating convergent and discriminant validity of temperament questionnaires for preschoolers, toddlers, and infants. *Developmental Psychology, 27,* 566-579.

Goldsmith, H. H., & Rothbart, M. K. (1991). Contemporary instruments for assessing early temperament by questionnaire and in the laboratory (pp. 249-272). In A. Angleitner & J. Strelau (Eds.), *Explorations in temperament: Contemporary conceptualizations, measurements and methodological issues.* New York: Plenum.

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., Hinde, R. A., & McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child Development, 58,* 505-529.

Hollingshead, A. (1975). *Four-Factor Index of Social Status.* Unpublished manuscript, Yale University, New Haven, CT.

Hubert, N. C., Wachs, T. D., Peters-Martin, P., & Gandour, M. J. (1982). The study of early temperament: Measurement and conceptual issues. *Child Development, 53,* 571-600.

Loehlin, J. C., Horn, J. M., & Willerman, L. (1990). Heredity, environment, and personality change: Evidence form the Texas Adoption Project. *Journal of Personality, 58,* 221-243.

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

Lytton, H. (1977). Do parents create, or respond to, differences in twins? *Developmental Psychology, 13,* 456-459.

Lytton, H., Martin, N. G., & Eaves, L. (1977). Environmental and genetical causes of variation in ethological aspects of behavior in two year old boys. *Social Biology, 24,* 200-211.

Matheny, A. P., Jr. (1980). Bayley's Infant Behavior Record: Behavioral components and twin analysis. *Child Development, 51,* 1157-1167.

Matheny, A. P., Jr. (1983). A longitudinal twin study of stability of components from Bayley's Infant Behavior Record. *Child Development, 54,* 356-360.

Matheny, A. P., Jr., & Brown, A. M. (1971). Activity, motor co-ordination and attention: Individual differences in twins. *Perceptual and Motor Skills, 32,* 151-158.

Matheny, A. P., Jr., Dolan, A., & Wilson, R. S. (1976). Twins: Within-pair similarity on Bayley's Infant Behavior Record. *The Journal of Genetic Psychology, 128,* 263-270.

Matheny, A. P., Jr., Wilson, R. S., & Dolan, A. (1976). Relations between twins' similarity of appearance and behavioral similarity: Testing an assumption. *Behavior Genetics, 6,* 343-351.

Matheny, A. P., Jr., Wilson, R. S., Dolan, A., & Krantz, J. Z. (1981). Behavioral contrasts in twinships: Stability and patterns of differences in childhood. *Child Development, 52,* 579-588.

McCall, R. B. (1977). Challenges to a science of developmental psychology. *Child Development, 48,* 333-344.

McCall, R. B. (1986). Issues of stability and continuity in temperament research. In R. Plomin & J. Dunn (Eds.), *The study of temperament: Changes, continuities and challenges.* (pp. 13-25). Hillsdale, N. J.: Erlbaum.

McCarthy, D. (1972). *Manual for the McCarthy Scales of Children's Abilities.* New York: The Psychological Corporation.

McCartney, K., Harris, M. J., & Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychological Bulletin, 107,* 226-237.

McDevitt, S. C. (1986). Continuity and discontinuity of temperament in infancy and early childhood: A psychometric perspective. In R. Plomin & J. Dunn (Eds.), *The study of temperament: Changes, continuities and challenges.* (pp. 27-38). Hillsdale, N. J.: Erlbaum.

McKeen, N. A., & Eaton, W. O. (1989). *Infant motor activity: Temperament behavior and sex differences.* Paper presented at the meeting of the Society for Research in Child Development. Kansas City, MO, April, 1989.

Neale, M. C., & Stevenson, J. (1989). Rater bias in the EASI temperament scales: A twin study. *Journal of Personality and Social Psychology, 56,* 446-455.

Nichols, R. C., & Bilbro, W. C., Jr. (1966). The diagnosis of twin zygosity. *Acta Genetica et Statistica Medica, 16,* 265-275.

Plomin, R. (1983). Developmental behavior genetics. *Child Development, 54,* 253-259.

Plomin, R. (1986a). Behavioral genetic methods. *Journal of Personality, 54,* 226-261.

Plomin, R. (1986b). Multivariate analysis and developmental behavioral genetics: Developmental change as well as continuity. *Behavior Genetics, 16,* 25-43.

Plomin, R., (1990). The role of inheritance in behavior. *Science, 248,* 183-188.

Plomin, R., Campos, J. J., Corley, R., Emde, R. N., Fulker, D. W., Kagan, J., Reznick, J. S., Robinson, J., Zahn-Waxler, D., & DeFries, J. C. (1990). Individual differences during the second year of life: The MacArthur Longitudinal Twin Study. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, and predictability,* pp. 431-455. Hillsdale, NJ: Erlbaum.

Plomin, R., Chipuer, H. M., & Loehlin, J. C. (1990). Behavior genetics and personality. In L. A. Pervin (Ed.), *Handbook of Personality Theory and Research* (pp.225-243). New York: Guilford.

Plomin, R., Coon, H., Carey, G., DeFries, J. C., & Fulker, D. (1991). Parent-offspring and sibling adoption analyses of parental ratings of temperament in infancy and early childhood. *Journal of Personality, 59,* 705-732.

Plomin, R., & DeFries, J. C. (1985). *Origins of individual differences in infancy: The Colorado Adoption Project.* Toronto: Academic Press Inc.

Plomin, R., DeFries, J. C., & Fulker, D. W. (1988). *Nature and nurture during infancy and early childhood.* New York: Cambridge University Press.

Plomin, R., DeFries, J. C., & McClearn, G. E. (1990). *Behavioral Genetics: A primer (2nd ed.).* San Francisco: W. H. Freeman & Co.

Plomin, R., Emde, R., Braungart, J. M., Campos, J., Corley, R., Fulker, D. W., Kagan, J., Reznick, S., Robinson, J., Zahn-Waxler, C., & DeFries, J. C. (1992). Genetic change and continuity from 14 to 20 months: The MacArthur Longitudinal Twin Study. Manuscript submitted for publication.

Plomin, R., & Foch, T. T. (1980). A twin study of objectively assessed personality in childhood. *Journal of Personality and Social Psychology, 39,* 680-688.

Plomin, R., & Nesselroade, J. R. (1990). Behavioral genetics and personality change. *Journal of Personality, 58,* 191-220.

Plomin, R., Pedersen, N. L., McClearn, G. E., Nesselroade, J. R., & Bergeman, C. S. (1988). EAS temperaments during the last half of the lifespan: Twins reared apart and twins reared together. *Psychology and Aging, 3,* 43-50.

Plomin, R., & Rowe, D. C. (1977). A twin study of temperament in young children. *The Journal of Psychology, 97,* 107-113.

Plomin, R., Willerman, L., & Loehlin, J. C. (1976). Resemblance in appearance and the equal environments assumption in twin studies of personality traits. *Behavior Genetics, 6,* 43-52.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52,* 569-578.

Rothbart, M. K. (1986). Longitudinal observation of infant temperament. *Developmental Psychology, 22,* 356-365.

Rowe, D. C., & Plomin, R. (1977). Temperament in early childhood. *Journal of Personality Assessment, 41,* 150-156.

Sattler, J. M. (1982). *Assessment of children's intelligence and special abilities* (2nd ed.). Boston: Allyn & Bacon, Inc.

Saudino, K. J., & Eaton, W. O. (1991). Infant temperament and genetics: An objective twin study of motor activity level. *Child Development, 62,* 1167-1174.

Scanlon, J. W. (1984). To ponder ponderal's length: A question to Lester et al. *Child Development, 55,* 669-671.

Scarr, S. (1966). Genetic factors in activity motivation. *Child Development, 37,* 663-673.

Scarr, S., & Carter-Saltzman, L. (1979). Twin method: Defense of a critical assumption. *Behavior Genetics, 9,* 527-542.

Scarr, S., McCartney, K. (1983). How people make their environments: A theory of genotype → environment effects. *Child Development, 54,* 424-435.

Stevenson, J., & Fielding, J. (1985). Ratings of temperament in families of young twins. *British Journal of Developmental Psychology, 3,* 143-152.

Thomas, A., & Chess, S. (1977). *Temperament and development.* New York: Bruner/Mazel.

Thompson, M. (1986). *Genetics in medicine* (4th ed.). Toronto: Saunders.

Thompson, L. A., Fulker, D. W., DeFries, J. C., & Plomin, R. (1988). Multivariate analysis of cognitive and temperament measures in 24-month-old adoptive and nonadoptive sibling pairs. *Journal of Personality and Individual Differences, 9,* 95-100.

Torgersen, A. M. (1981). Genetic factors in temperamental individuality. A longitudinal study of same-sexed twins from two months to six years of age. *Journal of the American Academy of Child Psychiatry, 20,* 702-711.

Torgersen, A. M. (1985). Temperamental differences in infants and 6-year-old children: A follow-up study of twins. In J. Strelau, F. H. Farley, & A. Gale (Eds.), *The biological basis of personality and behavior: Theories, measurement techniques, and development* (Vol. 1, pp. 227-239). Washington: Hemisphere.

Torgersen, A. M., & Kringlen, E. (1978). Genetic aspects of temperamental differences in infants. A study of same-sexed twins. *Journal of the American Academy of Child Psychiatry, 17,* 433-444.

Willerman, L. (1973). Activity level and hyperactivity in twins. *Child Development, 44,* 288-293.

Wilson, R. S. (1979a). Twin growth: Initial deficit, recovery, and trends in concordance from birth to nine years. *Annals of Human Biology, 6,* 205-220.

Wilson, R. S. (1979b). Analysis of longitudinal twin data: Basic model and applications to physical growth measures. *Acta Geneticae Medicae et Gemellologiae, 28,* 93-105.

Wilson, R. S. (1981). Synchronized developmental pathways for infant twins. In L. Gedda, P. Parisi, & W. Nance (Eds.), *Twin research 3: Intelligence, personality and development.* (pp. 199-209). New York: Alan R. Liss, Inc.

Wilson, R. S. (1983). The Louisville Twin Study: Developmental synchronies in behavior. *Child Development, 54,* 298-316.

Wilson, R. S. (1986). Continuity and change in cognitive ability profile. *Behavior Genetics, 16,* 45-60.

Wilson, R. S., & Harpring, E. B. (1972). Mental and motor development in infant twins. *Developmental Psychology, 7,* 277-287.

Wilson, R. S., & Matheny, A. P., Jr. (1986). Behavior genetics research in infant temperament: The Louisville twin study. In R. Plomin, & J. Dunn (Eds.), *The* *study of temperament: Changes, continuities and challenges,* (pp. 81-97). Hillsdale, N. J.: Erlbaum.

Wohlwill, J. F. (1973). *The study of behavioral development.* New York: Academic Press.

APPENDIX A

Subject Recruitment Letter

July 16, 1992

Dear Parent:

We hope all is well with you and Twin A and Twin B. Since we last wrote you with a description of our research findings we have been continuing our study and we now have a total 62 pairs of infant twins participating. Our results have been so encouraging that we have decided to begin a follow-up study of the children as toddlers, and we are writing to see if you would be interested in any further participation.

If you recall, we were interested in determining whether infant activity level had an inherited basis. Our initial results suggested that activity level was, in part, determined by heredity. However, because we only studied infants, we can't be sure that this is the case for older children. It is possible that when the children are older, environmental factors become a more important influence on motor activity. By studying the twins when they are toddlers, we should be able to gain a better understanding of the role that heredity plays in the development of activity level.

Like the initial study, the follow-up research would involve measuring the activity level of both twins in your own home. For two days, each child would wear the motion recorders while carrying on with their normal daily routines.

If you agree to participate in our study, we would arrange a convenient time for a home visit to bring the motion recorders. At this visit, we will conduct a quick assessment of motor development (general muscle co-ordination). We would also ask you to complete a brief questionnaires regarding the twins' activity levels and their degree of physical similarity. We would return two days later to collect the motion recorders and questionnaire and to weigh and measure the twins.

Parents who volunteer to participate will, once again, receive a summary of the results when the research is completed. As well, we will also provide a *revised* physical similarity diagnosis of their twins' zygosity. This should allow a more accurate indication of whether the twins are identical or fraternal.

If you definitely do not wish to participate please call 261-9075 and leave a message. If we do not receive a call in seven days, we will telephone you to answer any questions you may have and to see if you are interested in participating. If you do agree to participate, you would be free to withdraw from the study at any time.

Sincerely,

Warren O. Eaton, Ph.D. Kim Saudino, M.A.

APPENDIX B

Telephone Protocol

ID ___ Parent's name __________________ phone ______
Babies' name __________________ birthdate __________

Address ____________________________
____

Hello ... speak to parent... This is ..me... from the U of M psychology department...Calling..received letter about our follow-up study on infant activity...(No- see intro letter and explain)

Interested in participating? No OK that's fine. Thanks for time.

Yes or Not sure Perhaps tell.. more about study. We're following-up our initial research ... study... to see if there are any genetic influences on AL when the twins are toddlers... healthy .. need one parent at home with ...... is interested in .. volunteer. Study is much the same as before...measuring activity by means of special motion recorders .. like wristwatches .. that we bring your home .. get twins to wear for 2 days. Also do motor assessment and have parents complete a brief questionnaire regarding the twins' physical similarities.

If like participate .. I'll arrange convenient time to come .. bring actometers ..give you all details ..Also have questions re twins and family.

At end of 2 days .. come back .. 2nd visit .. collect actos diary. take Babies' weight length .. see how baby has developed as far as motor skills (see how baby is moving around, picking things up etc.)

Still interested? Correct name address pronunciation?

Spell twins' names ______________________

Directions ____________________________

Time appt ____________

my name + phone (488-8748 or 261-9075 for msg) in case sick

APPENDIX C

Consent to Participate

I, ____________________, along with my twins____________________, agree to participate in a research study of child activity behavior conducted by Kimberly Saudino and Dr. Warren O. Eaton, Department of Psychology, University of Manitoba. I understand that we are under no obligation to participate and that we may withdraw from the study at any time. I understand that information we provide for the study will be kept confidential to protect our privacy.

Date: ____________ Signature ____________________

Address to which results to be sent:

APPENDIX D

Demographic Questionnaire

General Information

ID ___ Date (YY/MM/DD) ________ Interviewer ___

Twins' names ________________________ Sex ___

Twins' age now _____

Have either of the twins had any health problems since the last visit, (i.e., six months of age)? ______

Specify ______________________________

____

How many children now live in your home? ____________

For each child: Sex (M,F) Birthdate (YY/MM/DD) Relation to baby (Full or Half sib,Unrelated)

	Sex (M,F)	Birthdate (YY/MM/DD)	Relation to baby (Full or Half sib,Unrelated)
1.	___	______	__________
2.	___	______	__________
3.	___	______	__________
4.	___	______	__________

Occupation and Education

We would like to ask some questions about you and your partner's occupational and educational backgrounds.

What is the highest grade or year (1 to 13) of secondary or elementary school ever attended?
mother ____________ father ____________

How many years of education have been completed at university?
mother ____________ father ____________

How many years of schooling have been completed at an institution <u>other than</u> a university, high school or elementary school? Include years of schooling at community colleges, institutes of technology, CEGEPS (general or professional), private trade schools or private business colleges, diploma schools of nursing etc.

mother ____________ father ____________

mother education classification __
father education classification __

Are you working now? Hours per week ____
If yes, what work do you do? __________________

Is your partner working now? (Y or N) ____
If yes, what work does he do? __________________
If no, did he work before? (Y or N) ____
If yes, what work did he do? ________________

mother occupation classification __
father occupation classification __

APPENDIX E

Actometer Instructions

A. Please leave the recorders on child as much as possible. It may be necessary to remove one or more of the recorders for dressing and undressing child. The recorders aren't waterproof, so be sure to remove them for baths. It is also very important for us to know of times when a recorder is off baby, so if you find it necessary to remove one or more of the recorders:

1) On the attached sheet note the time of day (not the time on the recorder itself) when each recorder is removed and re-attached.

2) Be sure to re-attach each recorder on the arm or leg from which it was removed. They are color-coded so you can check the attached sheet to see which recorder goes on which limb.

3) Be sure the recorder is snugly fastened on the outside of the wrist or ankle just above the wrist joint or ankle bone.

B. The recorders aren't fragile so you can treat your child as you normally do.

C. If we can not be present for the final recorder removal, we would like you to remove the recorders at the suggested time listed on the attached sheet (or as close to this time as practical). Record the actual time of removal and store the recorders in a place where they won't be disturbed until we can collect them.

If you are uncertain about what to do, please call:

Kim Saudino
488-8748
or
leave message at
261-9075

APPENDIX F

Actometer Record Sheet (Parent)

ID _____

A. Record of Motion Recorder Removals

Check the recorders which are removed

Right Arm	Left Arm	Right Leg	Left Leg	Time of day: Removed (hh:mm)	Time of day: Replaced (hh:mm)	Comments
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________
___	___	___	___	__ __ : __ __	__ __ : __ __	________

B. Final Recorder Removal

Best time to remove motion recorders: ___: ___ am/pm on __________.

Time of removal ___: ___ am/pm Date ____ ______ ______
(hh, mm; Day, Month, Year)

Once the recorders are removed, they should be moved as little as possible.

APPENDIX G

Instrument Measurement Properties

Actometers

Actometers are wrist watches that have been modified to record movement rather than time. In the present study, the Kaulins and Willis Model 101 Motion Recorder (actometer) were employed. This instrument has a conventional watch movement in which the hairspring and balance wheel have been removed. Movements of the watch case induce the pallet lever rocking motion which causes the hands to advance. Because the pallet lever is pendulous, it is responsive to accelerations on the case caused by tipping relative to gravity or by movements in the two spatial planes (horizontal and vertical) that are parallel to the watch face. Thus, when worn on the wrist or the ankle, the actometer will be responsive to the typical motor movements of the limb.

The Kaulins and Willis Model 101 Motion Recorder is a binary measure. All behaviors having a movement intensity above the instrument's threshold sensitivity will activate the actometer. Those behaviors with an intensity below the threshold will go undetected. Of those movements that are recorded, the intensity of motion is irrelevant; a vigorous movement will be treated in the same manner as a slow movement. Thus, using this device, activity is essentially dimensionalized in terms of above-threshold frequency of movement.

Actometer standardization

Each elapsed second on the Kaulins and Willis actometer is defined as one Activity Unit (AU). An AU reflects the number of movements required to advance the second hand from one marking on the dial to the next. Typically, the number of real minutes wearing-time is recorded; this allows for the calculation of an AU rate and enables comparisons across unequal wearing intervals.

Actometer reliability

The Kaulins and Willis actometer has been shown to be a highly reliable instrument for recording the activity of young children. Eaton et al. (1988) had 6-month-old infants wear four actometers, one per limb, for a 48-hour period. Using generalizability techniques, single limb actometer scores had an estimated reliability of .53 while the composited actometer scores had a reliability of .82. Similar estimates were found in the initial twin study (r_1 = .61, r_4 = .86).

One advantage of using a physical instrument, such as the actometer, to assess AL lies in the fact that it is possible to evaluate the reliability of the device independently of the human behavior it is designed to measure. This permits estimations of instrument generalizability and veridicality that are free from the confounding influences of behavioral variability. With this in mind, Eaton et al. (1988) used a chemical bath agitating machine to mechanically evaluate the reliability of the Kaulins and Willis actometer. Twenty-seven actometers were attached to a test-tube rack and agitated for 5-, 10-, and 15- minute trials. The results suggest a high degree of interchangeability among instruments; there were no significant differences among the readings of the actometers for any of the three trials and the intraclass correlation between measurement outcomes was .99. Moreover, the highly significant trials effect ($F[2,52] = 28{,}544.84$, $p < .0001$) indicated that actometer units increased with the amount of movement, thus demonstrating the veridicality of the measure.

Actometer validity

Correlations between the actometer and other methods of measuring AL have provided converging evidence for the construct validity of this standard system. Using the Kaulins and Willis Model 101 Motion Recorder, McKeen and Eaton (1989) found that 48-hour actometer-assessed AL in 6-month-old infants was significantly related to parent-rated activity on the Activity Subscale of Rothbart's (1981) IBQ ($r = .48$, $p < .001$). In addition, the amount of time the infants spent sleeping was *negatively* correlated with actometer units, thus providing evidence for discriminant validity.

Parent Activity Questionnaire

Infant Behavior Questionnaire

Parents in the initial study rated infant activity level with the Activity subscale of Rothbart's (1981) Infant Behavior Questionnaire (IBQ). The IBQ is a caregiver report instrument that is designed to assess the temperament of infants from 3 to 12 months of age.

Rather than requiring parents to make global judgments about their infants' behavior, the IBQ asks the parent to rate the approximate frequency of occurrence of concrete behaviors in various, specific situations encountered within the previous week (Goldsmith & Rothbart, 1991). In the present study, instructions were modified slightly so that ratings were based on the two-day interval in which the actometers were worn.

The Activity subscale comprises 17 items encompassing a broad range of daily activities (e.g., feeding, sleeping, playing). The parent is required to rate the infant's gross motor activity, "including movement of arms and legs, squirming, and locomotor activity" (Rothbart, 1981, p.573) in each situation. For example, "When being undressed, how often did your baby wave his/her arms and kick?" Responses range from 1 (never), to 4 (about half the time), to 7 (always), and to X (does not apply) when the parent has not seen the infant in the situation during the rating interval.

IBQ Activity subscale reliability. In their evaluation of the IBQ, Goldsmith and Rothbart (1991) present two kinds of reliability, "household reliability" and internal consistency. Household reliability, the correlation between mother ratings and the rating of a second adult in the household, was .69 ($p < .05$) for the activity scale. Internal consistency, as estimated by Cronbach's alpha, shows steady increases across three month intervals from 3 to 12 months of age. Goldsmith and Rothbart report coefficient alphas for the activity scale ranging from .73 for ratings of 3-month-old infants to .84 for ratings of 12-month-old infants. In the initial twin sample of 8-month-olds, alpha was estimated to be .88.

IBQ Activity subscale validity. The IBQ activity subscale has demonstrated substantial convergent validity. Goldsmith, Rieser-Danner, and Briggs (1991) found that, for both parents and teachers, ratings of activity on the IBQ and the Revised Infant Temperament Questionnaire [RITQ] (Bates, Freland, Lounsbury, 1979 cited in Goldsmith et al., 1991) were highly correlated ($r = .65$). Although at 3 months of age, there was no significant relation between IBQ Activity and actometer-assessed activity level (Eaton & Dureski, 1986), these measures have demonstrated moderate convergence at 6 months of age (McKeen & Eaton, 1989). Similarly, parent ratings of AL on the IBQ correlated .30 ($p < .01$), with the composite actometer score in the initial twin study.

Toddler Behavior Assessment Questionnaire

In the second phase of research, parents rated the activity level of their twins on the Activity subscale of the Toddler Behavior Assessment Questionnaire [TBAQ] (Goldsmith, 1987). The TBAQ scales correspond to those of the IBQ, and is considered to be the appropriate instrument to assess temperament when following up a sample previously assessed with the IBQ (Goldsmith, 1987; Goldsmith & Rothbart, 1991). As a measure of child temperament, the TBAQ is applicable to children ages 16- to 36- months of age. Like the IBQ, the TBAQ asks the parent to indicate the approximate frequencies of specific behaviors in specific situations during a designated time interval.

The Activity subscale consists of 20 items. The parent is requested to rate each child's "limb, trunk, or locomotor movement during a variety of daily situations, including free play, confinement, or quiet activities" (Goldsmith, 1987 p. 5). For example, "When playing inside how often did your child run through the house?".

TBAQ Activity subscale reliability. According to Goldsmith (1987), the internal consistency of the Activity subscale (as estimated by Cronbach's alpha) is .78. The average item-total correlation was .44, however, this is likely to be somewhat inflated by part-whole correlations (Goldsmith & Rothbart, 1991).

TBAQ Activity subscale validity. Goldsmith et al., (1991) reviewed the correlations between the TBAQ and other temperament questionnaires and have concluded that the Activity subscale of the TBAQ has shown considerable convergence with other rating measures of activity. Maternal reports of activity on the TBAQ correlated substantially with the activity scales of Buss and Plomin's (1975) EASI-III (r = .54) and Fullard, McDevitt and Carey's (1984, cited in Goldsmith et al., 1991) Toddler Temperament Scale (r = .73). Similar results were also apparent when activity level was rated by daycare teachers. Moreover, the intercorrelations between the TBAQ subscales suggests respectable discriminant validity.

Assessing Motor Development

Bayley Scales of Infant Development

Motor maturity in the initial phase of research was assessed using the Motor Scale from the Bayley Scales of Infant Development [BSID] (Bayley, 1969). The

BSID is, at present, considered to be the best available measure of infant development (Sattler, 1982). It is designed to evaluate an infant's developmental standing relative to other infants of the same chronological age. The age range of the scales is from 2- to 30- months. There are three components to the BSID, the Mental Scale (MDI); the Motor Scale (PDI); and the Infant Behavior Record (IBR).

The Motor Scale of the BSID is designed to provide a measure of the infant's degree of body control, co-ordination of large muscle groups and fine manipulative skills. The 81 items included in the scale are specifically directed towards behaviors reflecting fine and gross motor abilities such as, sitting, crawling, standing, walking, and grasping. From the Motor Scale, a standard score, the Psychomotor Developmental Index (PDI) is derived. The PDI is a normalized score with a mean of 100 and a standard deviation of 16.

BSID Motor Scale Standardization. According to Sattler (1982), the BSID is a well-standardized test. Both the Mental Scale and the Motor Scale were standardized on a sample of 1,262 normal, North American infants in fourteen age groups ranging from 2- to 30- months. Attempts to control for sex, race, residence and education of the head of the household were included in the standardization process. Generally, the sample is considered representation of the population; however, Bayley (1969) reports that there may be an underrepresentation of the rural population. The effects of this underrepresentation are deemed 'negligible' by Bayley.

BSID Motor Scale Reliability. Split-half reliability coefficients for the fourteen age groups range from .68 to .92 with a median of .84. The reliabilities for the first four age groups tend to be lower; however, at six months of age, the reliability coefficient is at a very acceptable level of .89.

McCarthy Scales of Childrens' Abilities

The *McCarthy Scales of Childrens' Abilities* [MSCA] (McCarthy, 1972) examines the abilities of children from 2½ through 8½ years of age. It contains 18 separate tests which are grouped into 6 scales: Verbal, Perceptual-Performance, Quantitative, General Cognitive, Memory and Motor. According to Sattler (1982), the MSCA is well standardized, psychometrically sound and shows promise for assessing the cognitive and motor abilities of children.

The Motor Scale of the MSCA is designed to assess the child's coordination in a variety of interesting and enjoyable, gross and fine motor tasks. It consists of 5 separate tests: 3 evaluating gross motor coordination (Leg Coordination, Arm Coordination, and Imitative Action); and 2 evaluating fine motor control as revealed by hand coordination and finger dexterity (Draw-A-Design, Draw-A-Child). From the Motor Scale the Motor Index, a standard score reflecting developmental level, is derived. The Motor Index has a mean of 50 and a standard deviation of 10.

MSCA Standardization. The MSCA was standardized on a sample of 1,032 children in the United States. There were approximately 100 children at each of 10 age levels (i.e., 2½, 3, 3½, 4, 4½, 5, 5½, 6½, 7½, 8½). The sample was stratified according to U.S. 1970 census data for the variables of sex, color, geographic region, residence (urban versus rural), and father's occupation. Overall, the sample is considered to be representative of the U.S. population.

MSCA Motor Scale Reliability. Split-half reliability coefficients of the Motor Scale for the 10 age levels range from .60 to .84 with a mean of .79. More specific to the present study, for ages 2½ and 3, the split-half reliabilities are .84 and .82 respectively. The stability coefficient of the Motor Scale, using a test-retest procedure with a 3 to 5 week interval, was .78 for ages 3 to 3½ years.

APPENDIX H

Summary of Derived Variables

Weight:
Mean of two measures in kg

Length:
Mean of two measures in cm

Head Circumference:
Mean of two measures in cm

Ponderal Index (PI):
(Weight × 1000) ÷ Length^3 (Scanlon, 1984)

Chronological Age in months (CA):
(Assessment Date - Birthdate) ÷ 30.4

Gestationally-adjusted Age in months (GA):
(Assessment Date - Mother-reported Due Date) ÷ 30.4

Interval in months:
(First Assessment Date - Second Assessment Date) ÷ 30.4

SES:
5 × [Mean of mother's and father's occupation rating] + 3 × [Mean of mother's and father's educational ratings] (Hollingshead, 1975)

Change Score:
(Time 2 score - Time 1 score)

Slope Score:
(Time 2 score - Time 1 score) ÷ Interval

Actometer Transformations:

AU	=	Stop time - Start time
MIN	=	Total minutes actometer wearing time
AU Rate	=	(AU ÷ MIN) × 30
AU Log	=	LOG10(AU Rate)

APPENDIX I

The TBAQ Activity Subscale (Goldsmith, 1987)

INSTRUCTIONS: Please read carefully before starting.

As you read each description of the child's behavior below, please indicate **how often** the child did this during the last **48 hours** by circling one of the numbers in the left column. These numbers indicate how often you observed the behavior described during the last **48 hours.**

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(NA)
Never	Very Rarely	Less than half the time	About half the time	More than half the time	Almost always	Always ----	Does not apply

The "Not Applicable" column (NA) is used when you did not see the child in the **situation** described during the last two days. For example, if the situation mentions the child going to the doctor and there was no time during the last 48 hours when the child went to the doctor, circle the (NA) column. "Does Not Apply" (NA) is different from "Never" (1). "Never" is used when you saw the child in the situation but the child never engaged in the behavior listed during the last 2 days. Please be sure to circle a number or NA for **every** item.

When playing inside (for example, because of bad whether), how often did your child:

1 2 3 4 5 6 7 NA Q1 run through the house?

1 2 3 4 5 6 7 NA Q2 climb over furniture?

When playing on a movable toy, such as a tricycle, how often did your child:

1 2 3 4 5 6 7 NA Q3 attempt to go as fast as s/he could?

When in a shopping mall or store, how often did your child:

1 2 3 4 5 6 7 NA Q4 seem eager to explore the store?

When your child joined in an active game with other children (for example, one that involved running or jumping), how often did s/he:

1 2 3 4 5 6 7 NA Q5 keep up with the more energetic and active children?

How often during the past two days did your child:

1 2 3 4 5 6 7 NA Q6 play games that involved running around, banging or dumping out toys?

1 2 3 4 5 6 7 NA Q7 play games that did not involve moving, such as looking at books or arranging toys?

When in the bathtub, how often did your child:

1 2 3 4 5 6 7 NA Q8 sit quietly?

1 2 3 4 5 6 7 NA Q9 splash or kick?

1 2 3 4 5 6 7 NA Q10 play with toys with a lot of energy? (if the child never has toys in the bath, mark "NA").

When being dressed or undressed, how often did your child:

1 2 3 4 5 6 7 NA Q11 squirm or try to get away?

1 2 3 4 5 6 7 NA Q12 lie or sit quietly long enough for you to get him/her ready?

When your child needed to sit still, as in church, a waiting room, or a restaurant, how often did s/he:

1 2 3 4 5 6 7 NA Q13 try to climb out of the chair?

1 2 3 4 5 6 7 NA Q14 play quietly with 1 or 2 toys?

1 2 3 4 5 6 7 NA Q15 try to climb all over other chairs?

When placed in a car seat or stroller how often did your child:

1 2 3 4 5 6 7 NA Q16 kick?

1 2 3 4 5 6 7 NA Q17 squirm?

1 2 3 4 5 6 7 NA Q18 sit still?

While a story was being read to your child, how often did s/he:

1 2 3 4 5 6 7 NA Q19 sit quietly?

1 2 3 4 5 6 7 NA Q20 get restless?

APPENDIX J

The Diagnosis of Zygosity

An essential first feature of any twin study is the diagnosis of zygosity. It is critical that zygosity determination be accurate because the value of twin research is dependent on the correct identification of MZ and DZ twins. Errors in diagnosis are conservative in terms of the genetic hypothesis. That is, they cause DZ intraclass correlations to be higher and MZ intraclass correlations to be lower than when diagnoses are correct. Hence, misclassifications of zygosity result in reducing estimates of genetic influence attained through the comparison of MZ and DZ intraclass correlations.

The most accurate methods of diagnosing zygosity is through DNA "fingerprinting" or a detailed blood analysis in which the members of each twin pair are compared on a number of genetic markers in the blood. Disconcordance on one or more markers would classify twins as fraternal. Although highly accurate, these techniques can be impractical in terms of the difficulty, expense, and ethical considerations involved in obtaining blood samples.

An alternative to blood typing is the diagnosis of zygosity based on physical similarity criteria. The use of questionnaires evaluating general and specific physical similarities between twins has been shown to be a valid method for determining zygosity. Nichols and Bilbro (1966) had teenagers complete self-report questionnaires reporting their hair color and texture, eye color, height, and weight; how they differed from their co-twin on these characteristics; and instances when their identity had been confused by parents, teachers, close friends, and acquaintances. Using the responses from twins whose zygosity had been previously determined by blood analysis, the authors developed objective two-stage decision rules to determine zygosity. Cross-validation of the diagnoses based on the decision rules with extensive blood typing indicated that there was and 87% correct classification from the physical similarity criteria. This rose to 93% when those cases that could not be classified from the rules alone, were diagnosed "intuitively" on the basis of all available information.

Cohen et al., (1973, 1975) extended the Nichols and Bilbro method by designing a questionnaire for parents that enabled the determination of twin zygosity during childhood. The brief 10-item questionnaire contains 6 questions regarding the degree of twin similarity for the physical characteristics of height, weight, facial appearance, hair color, eye color, and complexion; and 4 questions concerning general identity and instances of identity confusion. Using the responses to this questionnaire for children ages from 1- to 6-years who had been blood typed for zygosity, Cohen et al. performed a multivariate discriminant analysis to generate a set of discriminant function coefficients and a discriminant cutoff point which accurately classified over 90% of the twins. Those questions that contributed heavily to the discrimination of MZ and DZ twins included confusion by strangers and eye and hair color. By comparison, height and weight were the weakest discriminators.

In addition to being a valid method for determining zygosity, Cohen et al. (1975) have also demonstrated that this technique is reliable. In a test-retest study using their questionnaire, there was a correlation of .97 (p < .001) between the initial discriminant scores and the replication score taken 15 months later. Thus, it is clear that careful questioning that includes morphological information can provide reliable classifications of zygosity which have a high level of accuracy.

APPENDIX K

Zygosity Questionnaires

A. Parent Questionnaire

To what extent are your twins similar at this time for the following physical features (circle one):

	Not at all similar	Somewhat similar	Exactly similar
Height	0	1	2
Weight	0	1	2
Facial appearance	0	1	2
Hair color	0	1	2
Hair thickness	0	1	2
Hair curliness	0	1	2
Hair growth pattern	0	1	2
Amount of body hair	0	1	2
Eye color	0	1	2
Complexion	0	1	2
Ear lobe shape	0	1	2
Teething pattern	0	1	2

Q1. Do your twins look as alike as two peas in a pod?

Yes ___(1) No ___(0)

Q2. Do you or your spouse ever confuse the twins?

Yes, frequently ___(1) Occasionally ___(2) Rarely or never ___(0)

Q3. Are the twins sometimes confused by other family members?

Yes, frequently ___(1) Occasionally ___(2) Rarely or never ___(0)

Q4. Have close friends ever mistaken the twins?

Yes, frequently ___(1) Occasionally ___(2) Rarely or never ___(0)

Q5. Is it hard for strangers to tell the twins apart?

Yes, frequently ___(1) Occasionally ___(2) Rarely or never ___(0)

Q6. Do the twins differ in blood type?

Yes ___(1) No ___(0) Don't know ___(2)

Q7. Do twins run in your family?

Yes, mother's side ___(1) father's side ___(2) No ___(0)

Q8. Do you know whether the twins are identical or fraternal?

Yes, identical ___(1) Yes, fraternal ___(2) No ___(0)

Q9. If you do know whether they are identical or fraternal indicate how and by whom this was determined:

____

____

B. Researcher Questionnaire

To what extent are the twins similar at this time for the following physical features (circle one):

	Not at all similar	Somewhat similar	Exactly similar
Facial appearance	0	1	2
Hair color	0	1	2
Hair thickness	0	1	2
Hair curliness	0	1	2
Hair growth pattern	0	1	2
Amount of body hair	0	1	2
Eye color	0	1	2
Complexion	0	1	2
Ear lobe shape	0	1	2
Teething pattern	0	1	2

APPENDIX L

Zygosity Decision Rules (Adapted from Plomin & Rowe, 1977)

Twins are first evaluated according to birth and delivery information from hospital records. If any item applies, twins are diagnosed on that basis. When no diagnosis can be made from the hospital data, physical similarity criteria are then employed. Twins are observed with regard to the first level items. If any first-level item applies, the twins are diagnosed accordingly. When twins cannot be diagnosed at the first level, the twins are then examined on the second-level items. For each second-level item scored true, one point is assigned; the diagnosis with the larger number of points is then be regarded as true. If the total MZ and DZ points are equal, the twins cannot be classified objectively. For these cases, Nichols and Bilbro (1966) suggest that an intuitive diagnosis be made on the basis of all available information (level three).

Hospital data

Diagnosis of MZ if placenta is monochorionic.

Diagnosis of DZ if blood types differ.

Level 1

Diagnosis of DZ:

Distinctly different hair color or curliness.

Distinctly different eye color.

Distinctly different facial appearance.

Distinctly different skin complexion.

Twins never mistaken by casual friends.

Diagnosis of MZ:

Twins frequently mistaken by parents.

<u>Level 2</u>

One point towards diagnosis of DZ:

Slight differences in hair color, curliness or texture.

Slight differences in eye color.

Never mistaken by casual friends.

Difference in ponderal index.

Difference in head circumference.

(For both physical measures, differences were determined by taking the absolute value of the Twin A measure minus the Twin B measure. The within-pair differences were then rank ordered and split into three groups: 0 - least different; 1 - moderately different; 2 - most different. Pairs receiving a score of 2 were then considered to be different for the measure.)

One point towards diagnosis of MZ:

Occasionally or frequently mistaken by parents.

Occasionally or frequently mistaken by close friends or relatives.

Frequently mistaken by casual friends or acquaintances.

Level 3

At this level, we considered dental patterns, earlobe patterns, amount of body hair, fingerprints, and other information gleaned from the parent questionnaire.