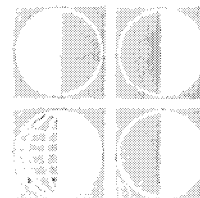


Environmental and Genetical Causes of Variation in Ethological Aspects of Behavior in Two-year-old Boys



*Hugh Lytton, †Nicholas G. Martin, and *Lindon Eaves

**Department of Educational Psychology, The University of Calgary, Alberta, Canada; and †Department of Genetics, University of Birmingham, Birmingham, England*

ABSTRACT: The interactive behavior of two-year-old male twins (17 MZ, 29 DZ pairs) with their parents was observed in the home and laboratory. Behavior counts, ratings, and experimental measures were subjected to a biometrical genetic analysis. A model which included only within- and between-families environmental effects fitted most variables best, except for Instrumental Independence (a rating) and Speech Rate (a count variable) which showed a significant genetic component. Low reliability may explain the small proportion of genetic variation in count variables, but the greater reliability of ratings and experimental measures (the latter, however, of doubtful construct validity) did not produce greater genetic determination. The largest part of the variance was explained by between-families environmental differences, only a minor part of which was due to interobserver differences. The probability of detecting genetic determination with this sample size, although it was large for an ethological study, is rather low.

Ethological methods and outlook have recently attracted behavioral scientists and have led them to look at human behavioral characteristics from an evolutionary perspective. Thus, attachment behavior has been viewed from the point of view of its evolutionary significance (e.g., Bowlby, 1971) and so has compliance with the norms of the social group (e.g., Stayton et al., 1971). An evolutionary perspective, however, inevitably implies partial genetical control of behavior, as evidenced by the question: "Why does this kind of animal solve those problems of survival in this way? (What is the evolutionary history of the behavior?)" (Blurton-Jones, 1974, p. 266). In the typical ethological study it is assumed, from a knowledge of the animal's ontogeny combined with observation, that learning plays little part in the development of certain kinds of behavior, which are then classified as innate. Both animal ethologists and those applying the ethological approach to human be-

havior have generally fought strangely shy of genetical studies proper (cf. Gould, 1974), although some exceptions exist, (e.g., Freedman, 1974).

Historically, ethological investigations have concentrated on describing "typical" patterns of behavior and have tended to ignore individual differences. Although this procedure may have been justified in the descriptive natural-history phase of the subject, an investigation of the causes of individual differences could lead to a more profound understanding of the biological importance and evolutionary history of the traits in question.

While it is possible to investigate human behavior and characteristics by means of questionnaires and tests, it is difficult to know how to approach the problem of social interactive behavior in the family or other group. It is in this area that the ethological approach has been increasingly used in recent years. When done with twins, such investigations provide a wel-

come opportunity to investigate the differences between individuals in observed interactive behavior from a genetical point of view.

This paper reports the biometrical genetical analysis of interactive behavior and other measures for two-year-old male sets of twins. The investigation was undertaken in order to assess the relative importance of genetical and environmental contributions to variation in certain social characteristics of young children, as well as to examine the development of these characteristics in the context of parent-child interaction. The characteristics focused on were compliance, attachment, independence, speech, and activity level. Child-parent interactive behavior was studied by means of naturalistic observation in the home, i.e., by an essentially ethological method, but was supplemented by interviews and ratings as well as by experimental procedures.

One assumption underlying the twin method, that parents make no greater distinctions between DZ twins than they do between MZ twins, is discussed (and essentially shown to be correct) in a separate paper (Lytton, 1977).

MATERIALS AND METHODS

SAMPLE

The sample consisted of 46 sets of male twins (17 MZ and 29 DZ), as well as 44 male singletons. Mean ages of the three groups were (in months): MZ, 32.0; DZ, 32.4; and singletons, 33.4. The age range was 25 to 35 months.

Twins were located through the birth registers of the local hospitals. The expectation is that after the first year of life surviving same-sex twin pairs represent 0.64 per cent of the population, and therefore 0.32 per cent will be same-sex pairs (Allen, 1955). This statistic provides an expectation of 25.6 same-sex male pairs

per year, given approximately 8,000 births per year in Calgary. We ascertained 25 born in 1969 and 26 born in 1970; therefore our ascertainment approximates the theoretical expectation very closely. Of these pairs, 14 were not included in the investigation: in three pairs, one partner had died, six families had left the city, two families did not speak English in the home, one could not be traced, and two refused to participate. While no data are available to compare those not included with those participating, the sample represents a very large proportion of the total male twin population born during those years, and there is no reason to suspect that the sample differs from the total population to any significant extent. Nine of the twin sets came from nearby towns. All available twins were seen, even if the mother was the only parent in the home (four families). The twin sample consisted of approximately one-third middle-class and two-thirds working-class families, based on father's occupation (Blishen, 1967), and singletons were chosen so that their social class distribution would be the same. The subjects were all white.

Zygoty of the twins was ascertained after the psychological investigation by blood typing of between 16 and 22 blood group systems. In five cases, when blood typing was impossible, zygoty was ascertained by finger ridge counts and physical similarity profile. By ascertaining zygoty after the testing, the chance was lessened of an observer having preconceived notions of how similar or different twins should be.

While many of the variables regressed on age, analysis of covariance showed that none of the differences in means of MZ and DZ groups and only three of the mean differences between twin and singleton groups (Variables 15, 16, and 23) could be accounted for by different age compositions of the groups.

Although every effort was made to obtain twin and singleton groups matched by father's occupation, when the sample had been ascertained it was found that groups differed slightly in the distribution of mother's education (see Table 1.) While the chi-squares for differences between twins and singletons and between MZ and DZ groups are not significant, it is evident that mothers' education is somewhat higher among singleton compared with twin families and among DZ compared with MZ families. (The latter could, of course, not be matched before the experiment.)

included in the observation whenever possible, but in some cases they were unavoidably absent. Therefore, we have records for fathers of only 38 sets of twins. The behavior of the target twins, as well as any behavior of others that impinged on them, was recorded continuously by dictating coded statements into a microphone held close to the mouth. The observation records were transferred to punched cards and analyzed by computer. A second observer was present for the second session in each home for the purpose of interviewing and rating the mother.

Mothers were interviewed regarding

TABLE 1
BREAKDOWN OF SAMPLE BY MOTHER'S EDUCATION AND TWINSHIP

Mother's Education Group	MZ Pairs	DZ Pairs	All Twin Pairs	Single- tons	Total
Not completed high school	9	9	18	11	29
High school graduate	5	10	15	16	31
Some college (incomplete or complete)	3	10	13	17	30
Total	17	29	46	44	90
Twins: Singletons $\chi^2(2) = 2.21$ N.S.					
MZ:DZ $\chi^2(2) = 2.47$ N.S.					

Differences between twins and singletons and between twin and singleton parents are discussed in a separate report (Lytton et al., 1977). In fact, only a few of these differences could be accounted for by disparities between the groups in mother's education.

PROCEDURE

Since the method employed in-the-home observations, and its attendant problems are discussed in detail in Lytton (1973), only a brief outline will be given here. Data on parent-child interaction were obtained from two three-hour observation sessions, during which unstructured family interaction was noted in predetermined code by trained observers. Fathers were

their perceptions of the children's characteristics as well as their own childrearing practices and attitudes, and some fathers were given a questionnaire containing parallel questions. Mothers were also asked to write an hour-by-hour account over a 24-hour period of all the incidents concerning the children and their own reactions to them ("diary").

MEASURES

Behavior Counts.—Specified types of behavior were categorized from the observation record, and summed frequency counts were computed for child, mother, and father. The measure of compliance (called "comply ratio") was the ratio of all instances of compliance to the total com-

pliant plus noncompliant behavior. Instances of certain kinds of behavior, e.g., attachment behavior, were converted to a "rate per minute." Frequencies of other types of behavior were expressed as a percentage of the total actions of the given agent (see Table 2.)

Peabody Picture Vocabulary Test and Experimental Measures.—The Peabody Picture Vocabulary Test was administered during an experimental session in a playroom. This session had as its main purpose the study of child compliance, attachment, and independence, experimen-

TABLE 2
MEANS, VARIANCES, AND INTRACLAS CORRELATIONS FOR MZ AND DZ GROUPS

NUMBER AND VARIABLE	TYPE*	MEANS			VARIANCES			INTRACLAS r	
		MZ	DZ	p	MZ	DZ	p	MZ	DZ
6 IQ-PPVT		78.27	87.93	0.003†	175.170	225.539	0.439	0.224	0.291
10 Comply ratio	C	0.66	0.62	0.059	0.010	0.009	0.713	0.619	0.444
18 Positive action	CR	0.63	0.62	0.851	0.050	0.032	0.138	0.903	0.848
19 Negative action	CR	0.71	0.77	0.305	0.073	0.061	0.543	0.510	0.532
14 Attachment	CR	0.15	0.15	0.880	0.005	0.003	0.089	0.600	0.714
30 Child speech	CR	1.24	1.52	0.012†	0.193	0.302	0.168	0.793	0.686
31 Child-mother speech ..	CR	0.55	0.74	0.034†	0.077	0.205	0.003	0.646	0.812
32 Child-father speech ..	CR	0.40	0.57	0.032†	0.068	0.146	0.032	0.741	0.684
86 Positive action	CP	0.17	0.17	0.901	0.002	0.001	0.021	0.900	0.800
87 Negative action	CP	0.19	0.21	0.178	0.004	0.003	0.336	0.500	0.429
92 Attachment	CP	3.87	4.01	0.653	3.328	1.662	0.021	0.556	0.586
28 Child command	CP	3.94	4.50	0.229	4.849	4.481	0.778	0.755	0.681
24 Activity shift	C	1.21	1.01	0.020†	0.196	0.126	0.142	0.906	0.867
25 Active behavior	CPT	33.45	31.91	0.387	71.360	64.597	0.727	0.878	0.811
26 Total activity	COM	49.26	46.28	0.085	77.533	53.382	0.214	0.894	0.850
9 Compliance	HR	2.78	3.03	0.199	0.821	0.841	0.960	0.675	0.645
13 Attachment	HR	2.74	2.74	1.000	0.401	0.356	0.680	0.434	0.510
16 Instrumental independence	HR	2.94	3.13	0.137	0.376	0.341	0.732	0.899	0.586
27 Speech maturity	HR	2.44	2.97	0.016†	0.709	1.122	0.158	0.858	0.910
17 Internalized standards	HR	3.02	2.84	0.343	0.654	0.523	0.518	0.782	0.642
8 Compliance	PR	3.97	4.17	0.189	0.511	0.380	0.353	0.276	0.198
12 Attachment	PR	2.79	2.56	0.191	0.496	0.555	0.760	0.192	0.267
15 Instrumental independence	PR	3.20	3.44	0.176	0.587	0.605	0.945	0.422	0.443
20 Toys	PR	20.19	20.35	0.934	57.125	89.383	0.190	0.333	0.118
21 Movement	PR	38.53	39.85	0.812	674.193	529.70	0.447	0.349	0.240
22 Total activity	PR	50.03	50.40	0.843	71.345	62.262	0.662	0.470	0.000
23 No. formboard pieces placed	PR	3.61	6.72	0.005†	17.581	21.407	0.593	0.560	0.713
11 Proximity	PR	2.90	2.64	0.261	1.320	0.679	0.054	0.591	0.200

* C, a count variable; CR, rate per minute; CP, percentage of child's actions; CPT, per cent of time; COM, composite of activity shift and active behavior, standardized score; HR, home rating; PR, playroom variable.

† Difference is nonsignificant after adjusting for regression on mother's education level.

Ratings.—Child compliance, attachment, independence, and speech maturity were also assessed by means of impressionistic ratings, based on observation, mother's "diary," and interview. These ratings were assigned by the observer who carried out two observations. The median interrater reliability for these home ratings was 0.67.

tally elicited by means of structured situations. The median interscorer reliability coefficient for the experimental measures was 0.82. Thus, from the standpoint of psychometric reliability, the experimental playroom measures are the most trustworthy.

All the 28 child variables used in the major analyses of this research are shown

in Table 2. Some of the variables are superordinate categories, representing the sum of several frequency counts, e.g., "attachment rate" sums instances of seeking mother's or father's proximity, attention, and help, "positive action" includes expressing affection, helping, showing pleasure, etc., "negative action" subsumes ignoring, yelling, refusing, hitting, etc.

INTERSITUATIONAL STABILITY AND INTEROBSERVER AGREEMENT

The stability of behavior over two observation sessions was assessed by calculating Spearman's ρ coefficients for the relative frequencies of some important behavior categories over all children. The coefficients for child behavior range from 0.67 for the rate of shift in activity to 0.21 for per cent of time spent in toy play, with a median of 0.49. The stability of behavior over two sessions is therefore only moderate, most of the coefficients being in the 0.40 to 0.50 range. It is, of course, not to be expected that children and parents engage in the same behavior to exactly the same extent on two occasions, since different situations will call forth different frequencies of given types of behavior. The data from two observation sessions were pooled in order to average out some of these fluctuations and to provide a more stable estimate of child or parent behavior.

Interobserver agreement was estimated by having two observers present in the home, each dictating a record (in a whisper) into a separate microphone. An agreement was counted when the same subject and predicate occurred for both observers within a thirty-second interval, and these agreements were summed over one-hour segments, taken from actual home observations. When actions recorded by one observer, but not by the other, were counted in with the disagreements, agreement reached 69.3 per cent,

and the median of four reliability checks was 63.9 per cent. When these actions were omitted from the calculations, maximum agreement was 85.6 per cent and the median 76.9 per cent, these figures being a measure of the extent to which two observers agree on the *coding* of behavior they both recorded (cf. Lytton, 1973, and Lytton and Zwirner, 1975, for further details).

The agreement between observers on the relative frequency overall of some important behavioral variables was also investigated. It was found that the two main observers agreed very closely on seven important variables ($p = 0.996$ for the null hypothesis of no difference by simultaneous confidence test, cf. Bock and Haggard, 1968). However, since four different observers were employed in the data collection, a certain amount of between-families variance could be contributed by differences between observers. This possibility will be examined below.

It is clear that the count variables of behavior in the home are rather unstable over time and subject to observer variation in recording. The likelihood of being able to detect reliable genetical or environmental variance against this background of behavioral lability within families and observer variation between families is correspondingly reduced. However, this argument is not applicable to the ratings and the experimental playroom measures.

In the case of five children, lack of cooperation during the Peabody Picture Vocabulary Test in the experimental playroom suggested that the results would be misleading, and the test was therefore repeated in the home a week or two later. We therefore have some information, though based on a very small sample and one of atypical cases, on the reliability of the PPVT. There were both upward and downward changes in IQ score, when the

test was given at home; they ranged from zero to 13 points, with an average of 5.6 points. The nonsignificance of Kendall's tau between the two occasions suggests some instability of the PPVT score at this age (judging by the present limited evidence) which might explain the peculiarities of this test's performance in the analysis (see below).

BIOMETRICAL-GENETICAL ANALYSIS OF THE DATA

FITTING GENOTYPE-ENVIRONMENT MODELS TO THE DATA

The assumptions and method required for the biometrical genetical analysis of twin data have been discussed in detail by Eaves and Eysenck (1975) and Jinks and Fulker (1970); only an outline will be given here.

The starting point for an analysis of the causes of variation is the between- and within-pairs mean squares obtained from an analysis of variance on pairs of twin scores:

	df	Expected mean squares
Between pairs	$n - 1$	$\sigma_w^2 + 2\sigma_b^2$
Within pairs	n	σ_w^2

The opportunity for partitioning genetical and environmental contributions to the total variance arises from the fact that the genetical expectations for σ_w^2 and σ_b^2 are different for MZ and DZ twins. We define E_1 to be a variance component due to individual environmental experiences within a twin pair (this will include chance environmental experiences and errors of measurement); E_2 is variance due to environmental differences between pairs and will include cultural and class differences; D_R represents additive genetical variance and is defined in detail in Mather and Jinks (1971), as are the other terms.

From the following equations,

$$\begin{aligned}\sigma_{bMZ}^2 &= E_2 + \frac{1}{2}D_R, \\ \sigma_{wMZ}^2 &= E_1, \\ \sigma_{bDZ}^2 &= E_2 + \frac{1}{4}D_R, \\ \sigma_{wDZ}^2 &= E_1 + \frac{1}{4}D_R,\end{aligned}$$

we can write our model for MZ and DZ mean squares in terms of the parameters and D_R in the following matrix form:

Mean square	E_1	E_2	D_R
MZ_b	1	2	1
MZ_w	1	0	0
DZ_b	1	2	$\frac{3}{4}$
DZ_w	1	0	$\frac{1}{4}$

We have four observed statistics (mean squares) and three unknown parameters for which we obtain estimates by the method of weighted least squares (see Eaves and Eysenck, 1975).

If we have s observed mean squares and estimate p parameters, then we are left with a residual sum of squares which, for large numbers, is distributed approximately as chi-square on $s-p$ degrees of freedom with which to test the fit of the model. We may fit other models which are subsets of the three parameter model given above, and below we fit models including the E_1 parameter alone, and two two-parameter models, E_1E_2 and E_1D_R .

If numbers are sufficiently large, the method of weighted least squares provides maximum likelihood estimates of genetical and environmental parameters and subsumes less efficient ratios such as those of Falconer (1960), Holzinger, and others which attempt to provide "heritability" estimates with no test of the implied model (Jinks and Fulker, 1970).

If either of the two parameter models fails, then there is justification for fitting a third parameter. What the third parameter might be, however, is largely a matter of choice in the absence of a further group of relatives, e.g., parents and offspring, adopted subjects, and so on. Any three-

parameter model which constrains the total variances to be equal will give identical residuals. Thus, although estimates are given for D_R , E_1 , and E_2 , we must recognize that the model may be reparameterized in a number of other more or less plausible ways. In this particular study, the problems are heightened by the small samples available, and in many variables it is impossible to discriminate between alternative models.

The general criteria to be satisfied, before fitting models in which no covariance between genotype and environment is assumed, have been stated by Jinks and Fulker (1970). Basically, one should see whether the MZ and DZ groups have been sampled from the same population by testing whether the total means and variances are equal, and one should do what one can to test for genotype times environment interaction. With a classic twin study such as this one, we are only able to test for interaction between within-pairs environmental effects (E_1 , measured by MZ absolute pair differences) and genotype (G) and/or between-families environmental effects (E_2 , measured by MZ pair sums).

Significant differences in means between the MZ and DZ groups were in fact found, as is shown in Table 2. However, these differences might possibly be explained by differences between the groups in mother's education, nonsignificant though the latter were. Multiple regression analyses were therefore performed, with the child variables as the dependent variables and mother's education and twin type (MZ vs. DZ) as the independent variables, entered in this order. The results showed that mother's education did indeed predict these variables. Once this fact had been allowed for, no significant difference between MZ and DZ groups remained in any variable, save Variable 23, and one such difference could be expected to occur by chance. In other words,

the original differences in these variables between the twin groups could be accounted for by differing distributions of mother's education (for further details, see Lytton et al., 1977). Table 2 also indicates where significant differences exist between the total variances of the MZ and DZ groups.

The variable showing one of the largest differences between groups is the Peabody Picture Vocabulary Test. The differences probably reflect inequalities of sampling and possible dependence of means and variances. The large day-to-day fluctuations in attention and motivation usually encountered in children of this age (note the seeming unreliability of the test reported above) would also have led to difficulties in standardization for this age range (cf. Lyman in Buros, 1965, p. 820). All this suggests that in this study the PPVT is a highly unsatisfactory measure and that little credence should be placed on the findings for this variable.

A test for $G \times E$ interaction in twin pairs (Jinks and Fulker, 1970), regression of absolute MZ pair differences on their pair sums, was also carried out for each variable. Genotype-environment interactions detected in this way produce distortions in the distribution of the variable which are a reflection of particular patterns of individual differences. Factors which produce linear regressions also tend to yield skewed distributions while kurtosis may reflect factors producing quadratic regression terms.

The worst distortions in scale are found in the count variables where the preponderance of very low scores, either rates per minute or percentages, produces strong positive skewness and often positive kurtosis as well. Consequently it is not surprising to find most of the sum-difference regressions among these variables. Significant (5 per cent or less) linear regression was found for Variables 6, 8, and 13;

significant quadratic regression for Variables 9, 10, and 18; both linear and quadratic regression for Variable 92.

It is well known that nonadditivity, which produces, for example, sum-difference regressions for twin pairs, can often be removed by transformation of scale (see Mather and Jinks, 1971). However, it was found in a previous analysis of twin-singleton differences (Lytton et al., 1977) and in a genetic analysis of questionnaire personality data (Martin and Eysenck, 1976) that transformation of scale often made very little difference to the results of the analysis of second-degree statistics in small samples. It was therefore decided to carry out the initial genetic analyses on untransformed data in the belief that the extra refinement would not substantially affect the conclusions. Inspection of the distributions of variables displaying sum-difference regressions (6, 8, 9, 10, 13, 18) suggested that a square-root transformation might remove the nonnormality of the distributions and the regressions. The transformation was successful for Variables 6, 9, and 18, but made little difference to the results of the genetical analysis of these variables.

RESULTS

No simple model will fit data that do not satisfy the criterion of equality of total variances for MZ and DZ groups. For this reason, variables in which the total MZ and DZ variables are different are omitted from the present genetical analysis. However, Eaves (1976) provides the foundation for an analysis of such differences. Individual variables that make up the linear composites "Total activity" for the home and for the playroom situation are also omitted, since the results are very similar to those of their composites. "Negative action" expressed as a per cent of total child actions is based on the same

data as when it is expressed as a rate per minute and has also been deleted.

The results of fitting the E_1 , E_1E_2 , E_1D_R , and $E_1E_2D_R$ models mentioned above to the remaining 18 variables are shown in Table 3. The table gives the residual chi-squares after fitting these four models by the method of weighted least squares. A model fits when the associated chi-square is not significant. The significance of the parameters \hat{E}_2 and \hat{D}_R from the fit of the $E_1E_2D_R$ model is also given.

The E_1 model tests the hypothesis that all the mean squares are the same, i.e., that all the variance is due to error or individual environmental experiences and none is caused by systematic cultural or genetical effects. We can see that this model fits the data for four of the five playroom experimental variables, and the data for PPVT IQ, a test that was administered to nearly all subjects in the playroom. This finding confirms our suspicions aroused by the poor construct validity of the experimental variables (discussed in Lytton, 1974) and problematic standardization of the PPVT. It suggests that much of the behavior elicited in the playroom was as arbitrary and artificial as the playroom situation itself. This finding should perhaps make us more wary of accepting results found in such experimental situations.

In the remainder of the variables, the E_1 model fails, implying that there are sources of variance over and above error and individual experience. The two-parameter models test respectively whether this extra variance can be adequately explained by between-families environmental effects (E_2) or additive genetical effects (D_R).

The E_1E_2 model fits every one of the variables and in only one case, Instrumental Independence Rating, does the E_1D_R model give a better fit. In many cases, the

TABLE 3
RESULTS OF MODEL FITTING FOR CHILD VARIABLES

Variable	Type*	$E_1(\chi^2)$	$p <$	$E_1 E_2 (\chi^2)^\dagger$	$E_1 D_R (\chi^2)$	$p <$	$E_1 E_2 D_R (\chi^2)$	\hat{E}_2 $p <$	\hat{D}_R $p <$	\hat{h}_2^2	$p <$
IQ-PPVT‡§		4.11	NS	0.60	1.02	NS	0.60	NS	—NS#
Comply ratio§	C	13.31	0.01	0.94	0.96	NS	0.23	NS	NS	0.32 ± 0.38	NS
Positive action§	CR	40.52	0.001	1.33	12.72	0.01	1.27	0.001	—NS
Negative action	CR	13.10	0.01	0.45	3.17	NS	0.27	0.05	—NS
Attachment	CR	21.88	0.001	3.45	8.07	0.05	1.40	0.001	—NS
Child speech‡	CR	26.05	0.001	3.50	4.84	NS	1.24	0.05	0.05	0.37 ± 0.21	0.05
Child command	CP	23.68	0.001	0.25	4.59	NS	0.06	0.01	NS	0.11 ± 0.24	NS
Total activity	COM	38.80	0.001	0.94	11.76	0.01	0.94	0.001	—NS
Compliance§	HR	19.96	0.001	0.06	3.98	NS	0.00	0.05	NS	0.07 ± 0.28	NS
Attachment§	HR	10.76	0.05	0.46	2.93	NS	0.14	0.05	—NS
Instrumental independence	HR	25.28	0.001	5.79	1.42	NS	0.09	NS	0.01	0.59 ± 0.23	0.01
Speech maturity‡	HR	41.24	0.001	1.22	17.90	0.001	1.22	0.001	—NS
Internalized standards	HR	17.84	0.001	0.63	3.07	NS	0.32	0.05	NS	0.17 ± 0.29	NS
Compliance§	PR	3.27	NS	0.84	1.02	NS	0.84	NS	NS	0.02 ± 0.60	NS
Attachment	PR	2.27	NS	0.13	0.58	NS	0.10	NS	—NS
Instrumental independence	PR	7.55	NS	0.01	1.48	NS	0.00	NS	—NS
Total activity	PR	4.32	NS	2.47	1.13	NS	0.20	NS	NS	0.90 ± 0.56	NS
No. formboard pieces placed‡	PR	17.30	0.001	0.60	5.77	NS	0.22	0.01	—NS

* C, a count variable; CR, rate per minute; CP, percentage of child's actions; CPT, per cent of time; HR, home rating; PR, playroom variable; COM, composite of activity shift and active behavior, standardized score.

† χ^2 nonsignificant for all variables under $E_1 E_2$ model.

‡ Indicates MZ and DZ means are significantly different (before correction for regression on mother's education).

§ Indicates significant regression of MZ absolute differences on pair sums.

|| Indicates that a square-root transformation removed the MZ sum-difference regression and that the results of model fitting to the transformed data were substantially the same.

Minus sign here and below indicates that a parameter takes a negative value.

latter model actually fails. The $E_1E_2D_R$ model also fits every variable, but only in the case of two variables—Instrumental Independence rating and Rate of Speech—does the $E_1E_2D_R$ model fit considerably better than the E_1E_2 model, and in these variables the genetic contribution plays a significant role. In all other cases it is non-significant, and in many instances even slightly negative. Where the D_R estimate is positive, a heritability estimate, $\hat{h}^2 = \frac{1}{2}\hat{D}_R/(\frac{1}{2}\hat{D}_R + \hat{E}_1 + \hat{E}_2)$ is given in the table and its significance level accords with that of D_R .

Thus, from our analysis of the twin data it appears that there is little genetical variance expressed in these measures of child behavior. All of the variance can be accounted for by error and individual experiences and by environmental differences between families, except for Instrumental Independence and Rate of Speech.

VARIANCE BETWEEN OBSERVERS

We have found that the E_1E_2 model seems the most appropriate for nearly all the measures of child behavior. Because of the low interobserver reliabilities, we have speculated that some of the variance between twin pairs (E_2) in the count variables can be traced to variation in the coding behavior of observers. Inter-rater agreement, however, was higher, and therefore this argument will not apply to ratings to the same extent and not at all to the experimental measures, which were all scored by the same experimenter.

To see how much of the environmental variance between families was due to observers, a hierarchical analysis of variance was carried out for all twin pairs, with observers ($N = 4$) as the first factor and twin pairs as the second factor, nested within observers. Since inspection of the distribution of mother's education between observers suggested differences between them as regards this measure,

mother's education was used as a covariate to correct for these differences.

The results are shown in Table 4 for those variables for which (a) \hat{D}_R was not significant and (b) the observer mean square was significantly larger (10 per cent level) than the pairs within observers' mean square. The table permits a comparison of the proportion of total variance due to observers (SS_O/SS_T), after covarying mother's education, and that due to between-families differences, $\hat{E}_2/(\hat{E}_1 + \hat{E}_2)$.

It will be noted that all the variables where a significant between-observers component is found are count variables. Child ratings, as expected, were less affected. Some of the differences originally noted between observers in the activity variables (Variables 24, 25 and 26) were evidently due to differences in mother's education, since taking out the effect of mother's education reduced the proportions of observer variance from their original higher levels. However, other variables were not affected by this.

In most of these variables, observer variance accounts for 10 to 20 per cent of the total, but it goes as high as 39 per cent for activity shift (Variable 24). A figure of 10 to 20 per cent of the total due to variance between observers seems reasonable for count variables. This finding suggests

TABLE 4
PROPORTION OF TOTAL VARIANCE
DUE TO VARIANCE BETWEEN
OBSERVERS AND TO \hat{E}_2

NUMBER AND VARIABLE	TYPE*	$SS_O/$ SS_T^\dagger	$E_2/$ $(\hat{E}_1 + \hat{E}_2)$
14 Attachment	CR	0.17	0.65
86 Positive action	CP	0.14	0.85
87 Negative action	CP	0.15	0.49
28 Child command	CP	0.17	0.71
24 Activity shift	C	0.39	0.88
26 Total activity	COM	0.21	0.87

* CR, rate per minute; CP, percentage of child's actions; C, a count variable; COM, composite of activity shift and active behavior, standardized score.

† Observer has been adjusted by covariance for mother's education.

that other causes account for the greater part of the between families variance, E_2 .

CONCLUSIONS

The only variables in this study for which the genetic contribution was significant were instrumental independence (a rating) and speech rate (a count variable). The latter variable which represents speech facility in a natural situation is probably a more reliable predictor of later intelligence than the PPVT score, a vocabulary score obtained in a test situation which depends on motivational and attentional factors that are very variable in two-year-old children.

The difficulty of measuring interactive behavior reliably by means of behavior counts has been noted. The low reliability over occasions and between observers and other anomalies were, no doubt, part of the reason why very little genetic determination of variation could be found in the count variables. But, although ratings and experimental scores showed greater reliability, the genetic contribution to these measures was no higher. (As discussed above, however, the construct validity of the experimental scores was doubtful.)

Furthermore, recent power calculations (Martin, in preparation) have shown that the probability of detecting even large amounts of genetic variation with samples of this size is rather low. Yet the authors know of no study involving detailed direct observation of behavior that has a larger sample, particularly since 44 singletons were also included in this investigation. This fact is not surprising in view of the expenditure of time involved. It took about 27 hours' work to gather and process all the data for one twin pair up to the time they were punched on cards. If such an ethological approach is thought useful because it increases ecological validity, one will have to be satisfied with what, for a genetical analysis, is a small sample.

The present results contrast with the findings of some earlier work, employing ratings (e.g., Scarr, 1969) and with those of some recent large twin studies (Eaves and Eysenck, 1974, 1975, 1976) which found intermediate heritabilities for a variety of adult personality dimensions assessed by self-report questionnaires. Furthermore, these studies suggested that environmental differences between families (E_2) did not contribute significantly to variation in these personality dimensions. On the other hand, the low degree of genetic determination detected in our investigation is consistent with some earlier research in personality (e.g., Gottesman, 1966, compared with Nichols, 1969). However, since environmental influences and gene expression may well change with age (e.g., Eaves and Eysenck, in press), comparisons between our study and those of adult subjects could well be misleading.

For most of the present personality characteristics, the largest part of the variance was explained by differences between families. These differences were partly due to interobserver differences, but they also reflected varying childrearing situations and differences in cultural milieu. The systematic differences between twins and singletons, for instance (see Lytton et al., 1977) seem to be an expression of such environmental variations. An interesting question is whether an investigation at a later age could detect greater genetic determination in variables collected in an ethological study such as this one.

ACKNOWLEDGMENTS

This research was supported by Canada Council Grant S71-1770. We are indebted to Pat Olsen, Arlene Barnes, Doreen Darby, Judy Eser, Laurel Saville, and Cecilia Schnurr who carried out the field work; to Marie Ellement of the Haematology Laboratory, Foothills Hospital, Calgary, who undertook the blood typing; and to Robert Black who assisted with the data analysis. We also wish to thank the Calgary hospitals for their cooperation.

REFERENCES

- ALLEN, G. 1955. Comments on the analysis of twin samples. *Acta Genet. Med. Gemell.* 4:143-159.
- BLISHEN, B. R. 1967. A socio-economic index for occupations in Canada. *Can. Rev. Sociol. Anthropol.* 4:41-53.
- BLURTON-JONES, N. 1974. Ethology and early socialization. In M. P. M. Richards (ed.), *The integration of a child into a social world*. Cambridge University Press, New York.
- BOCK, R. D., and E. A. HAGGARD. 1968. The use of multivariate analysis of variance in research. In D. K. Whitla (ed.), *Handbook of measurement and assessment in the behavioral sciences*. Addison-Wesley, Reading, Mass.
- BOWLBY, J. 1971. *Attachment and Loss, Vol. I: Attachment*. Penguin, London.
- BUROS, O. K. 1965. *The sixth mental measurements yearbook*. Gryphon Press, Highlands Park, N.J.
- EAVES, L. J. 1976. A model for sibling effects in man. *Heredity* 36:205-214.
- EAVES, L. J., and H. J. EYSENCK. 1974. Genetics and the development of social attitudes. *Nature* 249:288-289.
- . 1975. The nature of extraversion: A genetic analysis. *J. Pers. Soc. Psychol.* 32:102-112.
- . 1976. Genetic and environmental components of inconsistency and unrepeatability in twins' responses to a neuroticism questionnaire. *Behav. Genet.* 6:145-160.
- . Genotype and age interaction for neuroticism. *Behav. Genet.*, in press.
- FALCONER, D. E. 1960. *Introduction to quantitative genetics*. Ronald Press, New York.
- FREEDMAN, D. G. 1974. *Human infancy: An evolutionary perspective*. John Wiley, New York.
- GOTTESMAN, I. I. 1966. Genetic variance in adaptive personality traits. *J. Child Psychol. Psychiat.* 7:199-308.
- GOULD, J. L. 1974. Genetics and molecular ethology. *Zeitschrift f. Tier-psychol.* 36:267-292.
- JINKS, J. L., and D. W. FULKER. 1970. Comparison of the biometrical, MAVA, and classical approaches to the analysis of human behavior. *Psychol. Bull.* 73:311-349.
- LYTTON, H. 1973. Three approaches to the study of parent-child interaction: ethological, interview and experimental. *J. Child Psychol. Psychiat.* 14:1-17.
- . 1974. Comparative yield of three data sources in the study of parent-child interaction. *Merrill-Palmer Quart.* 20:53-64.
- . 1977. Do parents create, or respond to, differences in twins? *Devel. Psychol.* 13:456-459.
- LYTTON, H., D. CONWAY, and R. SAUVÉ. 1977. The impact of twinship on parent-child interaction. *J. Pers. Soc. Psychol.* 35:97-107.
- LYTTON, H., and W. ZWIRNER. 1975. Compliance and its controlling stimuli observed in a natural setting. *Devel. Psychol.* 11:769-779.
- MARTIN, N. G., and H. J. EYSENCK. 1976. The genetics of sexual behavior. In H. J. Eysenck, *Sex and personality*, Chap. 6. Open Books, London.
- MATHER, K., and J. L. JINKS. 1971. *Biometrical genetics*. Chapman and Hall, London.
- NICHOLS, R. C. 1969. The resemblance of twins in personality and interests. In M. Manosevitz, G. Lindzey, and D. D. Thiessen (eds.), *Behavioral genetics: Methods and research*. Appleton-Century, New York.
- SCARR, S. 1969. Social introversion-extraversion as a heritable response. *Child Devel.* 40:823-832.
- STAYTON, D. J., R. HOGAN, and M. D. S. AINSWORTH. 1971. Infant obedience and maternal behavior: The origins of socialization reconsidered. *Child Devel.* 42:1057-1070.