

# Dimensions of Psychological Masculinity-Femininity in Adult Twins from Opposite-Sex and Same-Sex Pairs

John C. Loehlin<sup>1,3</sup> and Nicholas G. Martin<sup>2</sup>

Received 8 Feb. 1999—Final 5 Nov. 1999

Male and female twins with opposite-sex co-twins were compared to twins with same-sex co-twins on three independent dimensions of masculinity-femininity, in order to examine the hypothesis that the hormones of the co-twin might have an effect on prenatal masculinization. The analysis was originally carried out for an older cohort from the Australian Twin Registry (2647 pairs, mean age 41.2), and then repeated in a younger cohort (1503 pairs, mean age 23.2). For women, the results on two of the three scales support and extend that of an earlier large study in Finland by Rose *et al.* (1994), who found no effect of sex of co-twin on feminine interests. One of the two scales contrasted worried and calm individuals, the other, confiding and reserved ones. The third scale, willingness to break or bend rules, showed a small effect of shared environmental influence, and it lay in the expected direction for a prenatal hormonal effect—females with a male co-twin scored higher (more like males). Most previous studies have not looked at the effect of sex of co-twin on males. The present study detected several such effects, although all were small in magnitude. The pattern was complex: sometimes the effect was in the masculine direction, sometimes in the feminine direction; sometimes there was agreement between the older and younger cohorts, sometimes not. Overall, the results suggest that no simple masculinization hypothesis—prenatal or postnatal—will adequately account for all the evidence. Age, sex, and aspect of masculinity-femininity must be taken into account.

**KEY WORDS:** Opposite-sex twins; masculinity-femininity; prenatal masculinization; hormones.

## INTRODUCTION

There has been interest in recent years in comparing female twins who have opposite-sex and same-sex co-twins for possible evidences of prenatal masculinization by the hormones of a male co-twin. The initial impetus for such studies was the finding of such masculinizing effects in litter-bearing mammals. Females that had been located between two male fetuses showed measurable differences in a number of relevant anatomical and behavioral characteristics, compared to females that had been located between two female fetuses (e.g., vom Saal, 1989). The presumption was that

androgens from developing male fetuses diffused through the fetal membranes and affected the development of adjacent female fetuses in a masculine direction. In some animal studies, the maternal circulation has also been implicated in hormone transfer (for a discussion, see Miller, 1994). An extensive review of how prenatal hormones may affect behavior is presented by Collaer and Hines (1995).

The evidence from humans has been mixed. Cole-Harding *et al.* (1988) reported that females from opposite-sex pairs scored higher on a mental rotations test than did females from same-sex pairs. Males typically average higher than females on this test. Resnick *et al.* (1993) reported that females from opposite-sex fraternal twin pairs averaged higher (more like males) on a questionnaire measure of sensation-seeking than did females from same-sex pairs. Miller (1994) found females from opposite-sex pairs to have more

<sup>1</sup> Department of Psychology, University of Texas, Austin, Texas 78712. Fax: 512-471-6175. e-mail: loehlin@psy.utexas.edu.

<sup>2</sup> Queensland Institute of Medical Research, Herston, Brisbane, Australia.

<sup>3</sup> To whom correspondence should be addressed.

masculine attitudes than females from same-sex pairs. (He looked at males as well, and found smaller effects, but also tending in the direction of a shift toward the opposite sex). The differences in these studies might reflect prenatal hormones, but, as the authors noted, they might reflect postnatal effects as well: girls growing up with twin brothers might well differ in their attitudes and experiences from girls growing up with twin sisters.

A trait on which a postnatal interpretation is less plausible is otoacoustic emissions (weak sounds emitted by the inner ear). McFadden (1993) reported that a small sample of females from opposite-sex twin pairs averaged fewer such emissions (i.e., were more male-like), than females from same-sex twin pairs. Postnatal factors seem less plausible in this case, as characteristic sex differences in otoacoustic emissions are evident in infants at one month of age (Burns *et al.*, 1992), and a trait that almost nobody is aware that they possess is unlikely to be directly socialized.

However, a number of studies on other traits that differ between the sexes have yielded negative findings. Henderson and Berenbaum (1997) looked at sex-typed play in 71 three- to eight-year-old girls who were members of same-sex or opposite-sex dizygotic twin pairs. They found no difference in preference for playing with sex-typed toys for girls with same-sex or opposite-sex co-twins, either during direct observation or by mother's report. (Sex-typed play included, for example, playing with dolls or toy dishes versus playing with toy cars or trucks.) Rodgers *et al.* (1998), in a study of 83 seven- to twelve-year-old twin pairs also found no difference in gender-typed play for either boys or girls having same- or opposite-sex co-twins. Rose *et al.* (1994) found no difference on a scale of feminine interests among 16-year-old girls from opposite-sex and same-sex pairs in a large Finnish twin sample. They also found no difference in the age at which the first child was born to women from the two kinds of twin pairs. Loehlin and Martin (1998) examined some 90 variables related to reproduction and childrearing among adult female twins from a large Australian sample, and concluded that if there were any differences on such variables between twins who had or had not been prenatally exposed to hormones from male co-twins, they must be very small in magnitude.

The present study, also using the Australian Twin Registry, looks at three dimensions of masculinity-femininity in females having male or female co-twins. These are scales derived in an earlier study (Loehlin *et al.*, in press) from items of two personality inventories. They represent within-sex variation along three

different dimensions on which males differ from females: the extent to which they admit to fears and worries, the extent of reserving their feelings from others, the extent to which they are willing to bend or break social rules. These three scales do not, of course, exhaust the ways in which males and females differ in personality and attitudes: for example, there are substantial between-sex differences as well as within-sex variation on measures of interests and vocational preferences (Lippa, 1995). Nevertheless, these scales do extend the dimensions of sex differences on which the effects of prenatal masculinization might be observed.

The present study also includes males. Although most prior studies have focused on prenatal effects on females, based on the research with lower animals, it remains, of course, a question of some interest whether the hormones of the prenatal partner might have a direct or indirect effect on the degree of masculinization in males as well.

Another feature of the present study is to examine the heritability of the three dimensions involved. We begin by asking whether the genes underlying individual differences on the trait are the same in males and females. If they are not, it makes it less likely that a common mechanism accounts for variation within the two sexes. In addition, heritability analyses can play an important role in deciding whether it is differences in exposure to a condition (such as the level of a hormone) that are responsible for variation in a given trait, or differences in sensitivity to the exposure. If the major influence is pair-to-pair differences in exposure, this should raise all twin correlations, and be reflected in a substantial estimate for the shared environmental parameter ( $c^2$ ). On the other hand, if gene-based differences in sensitivity are responsible, these will differ for dizygotic but not monozygotic pairs, and will contribute to heritability ( $h^2$ ) but not  $c^2$ .

In a second study, based on a later, younger Australian twin cohort, the generality of the findings of the first study are assessed.

## METHOD

### Measures

In a previous study (Loehlin *et al.*, in press), three masculinity-femininity scales were derived from the items of two personality inventories that had been given to roughly 3000 adult Australian twins. The inventories were short versions of the Eysenck Personality Questionnaire (Eysenck *et al.*, 1985) and the Cloninger

Tridimensional Personality Questionnaire (Cloninger *et al.*, 1991). The derived masculinity-femininity scales contained 12, 6, and 6 items respectively; they were approximately orthogonal within both the male and female subsamples, and the pattern of item factor loadings agreed well for the two sexes.

The first of the scales contrasts individuals who describe themselves as fearful and worried with those who describe themselves as calm and confident. It will be referred to as “Worried.” Women averaged about two-thirds of a standard deviation higher than men on this scale. The second scale, “Reserved,” contrasts individuals who tend to keep their feelings to themselves with those who discuss them freely with others. Men averaged higher than women on this scale, by about one-half a standard deviation. The third scale, “Breaks rules,” contrasts individuals who are willing to bend or break social rules when it is to their advantage with individuals who conform more strictly to convention. Men averaged higher than women on this scale, again by about a half standard deviation. In the earlier study, the second scale, Reserved, showed a low but significant correlation with the number of repeats of a CAG triplet in the androgen receptor gene, whereas the first and third did not; however, it is clear that at most only a small fraction of the variation of the trait is associated with this particular gene.

### Model Fitting

In the present study we examine first the extent to which each of these three masculinity-femininity scales has the same genetic basis in males and females, by fitting structural equation models to means and covariances in five groups: monozygotic (MZ) male pairs, MZ female pairs, dizygotic (DZ) male pairs, DZ female pairs, and opposite-sex DZ pairs. The model used is the general model for sex limitation described in Neale and Cardon (1992, pp. 212–217). We follow their method, which is to estimate separately in one sex the genetic variation common to both sexes and that distinctive to the sex in question. To the extent that the latter is small, we conclude that the same genes affect the trait in both sexes. The analysis also provides estimates of the heritabilities of the traits for males and for females.

Next, we compare the means on the three masculinity-femininity scales for females and males having opposite-sex and same-sex co-twins. A model-fitting procedure incorporating age is used, since there are some average age differences among the groups.

Finally, the above analyses are repeated for twins from a younger cohort from the same twin registry.

### Participants

*Older Cohort.* In 1988, 3808 pairs of twins from the Australian Twin Registry who had participated in a 1980 mail study (Jardine *et al.*, 1984) were sent a lengthy follow-up questionnaire on various health-related topics. Included were the two personality inventories whose items were used in deriving the masculinity-femininity scales. In 2647 pairs (70%) both twins provided scores on the three scales. Females and MZ twins were more heavily represented, as in most volunteer twin samples (Lykken *et al.*, 1978). The sex/zygosity subgroup sizes ranged from 930 to 223 pairs, and the twins' ages ranged from 24 to 87 years, with a mean of 41.2 years.

*Reliability Study.* In 1990, after a 2-year interval, 500 males and 500 females who had responded within the first month in the main study were sent the same questionnaire again. Scores on the three masculinity-femininity scales on both occasions were obtained for 872 of these individuals, and used to calculate test-retest reliabilities.

*Younger Cohort.* In 1990, an additional group of twins from the Australian Twin Registry who had been too young for selection into the previous studies was sent a questionnaire which included the inventories on which the three masculinity-femininity scales were based. The 3269 contactable pairs in this cohort were born between 1964 and 1971, and were 17–28 years of age at the time of testing. Scores were available for both twins for 1503 pairs (46%), again more heavily weighted toward females and MZ twins, with sex/zygosity subgroup sizes ranging from 454 to 162 pairs. The lower response rate probably reflects the younger ages of this group, and the fact that they were not recruited from among twins who had already participated in a similar study.

### RESULTS—OLDER COHORT

Correlations between twin pair members on the three masculinity-femininity scales are shown in Table I. In general, correlations are higher for MZ than for DZ pairs, suggesting that the genes contribute to individual differences on each of the three measures. With one or two exceptions, correlations for opposite-sex DZ pairs are about the same size as those for same-sex DZ pairs, suggesting that the same genes are affecting the traits in both sexes. In interpreting these correlations,

**Table I.** Older Cohort: Correlations of Twins on Three Masculinity-Femininity Scales in Five Groups, Mean Ages, and Reliabilities of the Scales<sup>a</sup>

Group	Worried	Reserved	Breaks rules	Mean age	Ns
MZf	.46	.35	.39	42.6	930
MZm	.40	.35	.30	40.8	398
DZf	.23	.14	.31	42.0	531
DZm	-.05	.12	.17	38.5	223
DZmf	.08	.11	.19	39.8	565
Reliabilities					
Alpha—f	.80	.69	.63		1860
Alpha—m	.81	.70	.66		883
Test-retest	.80	.71	.74		872

<sup>a</sup> MZ = monozygotic, DZ = dizygotic, f = female, m = male. Pearson correlations between first and second twin. Ns for correlations are number of pairs, for reliabilities are number of individuals. Alphas are Cronbach alpha internal-consistency reliability coefficients, based on one twin from each pair. Test-retest is based on a sample retested after 2 years.

one should keep in mind that they are depressed by the less-than-perfect reliability of these rather short scales. Alpha and test-retest reliability coefficients are given in the last three rows of Table I. The two types of reliability are in good agreement. Most of the coefficients fall in the .60 to .80 range, and the reliabilities are quite similar for males and females. Mean ages of the various groups are also shown in the table. Note that the female twins average a couple of years older than the males. The opposite-sex pairs resemble the males. Apparently, the age of the male twin was the critical factor in selecting the pair into the study.

The general impressions about heritability derived from inspection of the correlations were supported by formal model-fitting analyses using covariances and means in the five twin groups, and incorporating age. As a first step, sex-limitation models were fit separately for the three traits. Apart from age, these are basically three-parameter heredity-environment models fit separately for males and females, except that in one sex (in this case males) the additive genetic component is split into a portion shared with the other sex and a portion unique to the sex in question. If the unique portion is negligible, it suggests that the same genes influence the trait in both sexes. This was evaluated by a chi-square difference test between the full model and a model without the extra component. Three basic heredity-environment models were considered for each trait, and the best-fitting one used for this and subsequent analyses. All three models had the same first and third

parameters: additive genetic effects,  $a$ , and a residual,  $e$ , comprising unshared environmental effects, interactions, errors of measurement, and so forth. However, they differed in the second parameter, which could either be  $c$ , shared environment,  $d$ , genetic dominance, or  $i$ , multiple-gene epistasis. An *aie* model fit best for the first two scales, Worried and Reserved, and an *ace* model fit best for the third, Breaks rules. Results from fitting these models are shown in Table II.

The first two rows of Table II show the  $\chi^2$ s for the full sex-limitation model (28df) and the model fixing the sex-specific genetic effect to zero (29df). Clearly, the fit is trivially, if at all, improved by allowing for sex limitation, and we will henceforth use the simpler model that assumes the same genes act in both sexes, although possibly to different degrees. The next two lines of the table show that the fit of this model is excellent. The root mean square error of approximation (RMSEA) estimates the fit of the model in the population, after allowing for sampling error. A common rule of thumb (Steiger, 1989; Browne & Cudeck, 1993) is that values of RMSEA below .05 constitute excellent fits. As indicated in the third row of the table, even the upper 90% confidence limit for this parameter falls below .05 for all three traits. (The estimated value of RMSEA is zero for Worried and Reserved, based on a chi square less than its degrees of freedom, and is close to zero for Breaks rules.) The standardized root mean square residual (SRMR) reflects the average discrepancy between the observed data and that implied by the fitted model. A value below .08 may be considered to represent a relatively good fit (Hu & Bentler, 1999). Clearly, by either criterion the models fit well for all three traits.

The next section of Table II gives the parameters for the heredity-environment models and regressions with age, and the bottom section shows broad heritabilities calculated from the heredity-environment parameters. On the face of it, the heredity-environment parameter estimates for Worried and Breaks rules appear quite different across the sexes, but those for Reserved are similar for males and females. These conclusions are supported by the chi square difference tests in Table III, which are highly significant for the former two, but trivial for the latter. The differences appear mostly to reflect differences in the second parameter ( $i$  for Worried,  $c$  for Breaks rules). The broad heritabilities (bottom of Table II) show less difference across the sexes, although those for the males run a bit lower for all three traits. The heritabilities tend to lie in the .3s or .4s for all of the traits, with Breaks rules slightly

**Table II.** Older Cohort: Fit Statistics, Parameter Estimates, and Heritabilities<sup>a</sup>

Parameter	Worried	Reserved	Breaks rules
<b>Fit statistics</b>			
$\chi^2$ (28 <i>df</i> )	21.30	25.03	29.28
$\chi^2$ (29 <i>df</i> )	21.30	25.08	29.28
RMSEA-ul90%	.018	.026	.034
SRMR	.022	.038	.040
<b>Parameter estimates</b>			
$a_f$	.68	.45	.42
$i_f$	.03	.35	.0
$c_f$	.0	.0	.41
$e_f$	.74	.80	.78
$a_m$	.19	.44	.50
$i_m$	.61	.38	.0
$c_m$	.0	.0	.05
$e_m$	.77	.81	.82
$r_{age,m}$	-.05	.17	-.22
$r_{age,f}$	.04	.10	-.25
<b>Heritabilities</b>			
$h^2_f$	.46	.43	.36
$h^2_m$	.41	.38	.27
$h^2_{f,corr}$	.58	.62	.57
$h^2_{m,corr}$	.49	.54	.41

<sup>a</sup> Parameters: a = additive genetic, i = epistatic genetic, c = shared environment, e = residual (unshared environment, error, interaction, etc.);  $r_{age}$  = fitted within-group correlation of trait with age. .0 = parameter fixed to zero. Subscripts: f = female, m = male, corr = corrected. RMSEA-ul90% = upper limit of 90% confidence interval for root mean squared error of approximation; SRMR = standardized root mean squared residual,  $h^2$  = heritability.

**Table III.** Older Cohort: Chi-Square Tests of Parameter Differences Between Sexes

Test	<i>df</i>	Worried	Reserved	Breaks rules
Same a, i/c, e	3	16.88	.32	50.96
<i>p</i>		.0007	.96	<.0001
Same $r_{age}$	1	7.68	9.57	6.31
<i>p</i>		.006	.002	.01

lower. The shared environment parameter  $c^2$  is absent for the first two traits, and negligible for males for Breaks rules, but is modestly present ( $c^2 = .41^2 = .17$ ) for females for the latter.

These traits are, of course, measured with error. The last two rows of Table II show the heritabilities after correction for unreliability by dividing them by their respective alphas from Table I. Most now fall around .50, a value fairly typical for the heritability of personality measures. The female heritabilities remain

a little higher. Reliability correction would raise the  $c^2$  value for females on Breaks rules from .17 to about .27.

The correlations with age are significantly different across the sexes (Table III) for all three traits, although as the Table II values indicate, the differences from trait to trait are much more striking. Individuals of both sexes tended to be more reserved and to break rules less as they grew older—males slightly more so in the first case, and females in the second. The overall age correlations for Worried are small—the men grew a little less anxious with age, and the women a little more anxious.

Table IV provides the data of main interest. It compares twins who have opposite-sex and same-sex co-twins. The results for female twins are given in the top part of the table. The differences between the means of those having opposite-sex and same-sex co-twins are small in size, and only one, Breaks rules, is even in the expected (masculine) direction. None of the differences is even close to being statistically significant: the probabilities of differences of this magnitude occurring by chance in sampling are all above .5. The effect sizes are well below the .20 level conventionally considered “small” (Cohen, 1977), and are much smaller than the effect sizes due to gender itself (.48 to .66, bottom of the table). The hormones of a male co-twin apparently have not markedly affected any of these aspects of the femininity of these women.

What about the males? The second part of Table IV tells the story, which is a different one. The males with female co-twins average significantly higher on two of the scales, Worried and Breaks rules, although the effect sizes are still small, either absolutely or in comparison with the differences between the sexes. Note that these differences are in opposite directions with respect to femininity—having a female co-twin apparently makes a man more feminine in his anxieties and less feminine in his conscience—which suggests that it cannot be a simple matter of being feminized by exposure to female hormones *in utero*.

**RESULTS—YOUNGER COHORT**

It will be recalled that corresponding data were available for a later cohort of twins from the same twin registry. These twins were both younger and more restricted in age range than those in the older cohort. Their mean age was 23.2 years with a standard deviation of 2.2 years, compared to 41.2 and 12.8, respectively, in the older cohort.

Table V shows twin pair correlations, age means, and *N*s in the various groups. As in Table I, there tend

**Table IV.** Older Cohort: Means on Three Masculinity-Femininity Measures for Twins with Opposite-Sex and Same-Sex Co-Twins<sup>a</sup>

Group	Worried	Reserved	Breaks rules	N
Females with				
OS co-twin	26.51	11.31	9.37	691–692
SS co-twin	26.32	11.37	9.14	3158–3170
$\chi^2_{\text{diff}}$ (1 df)	.11	.04	.17	
<i>p</i>	.74	.84	.68	
effect size	-.03	-.02	.08	
Males with				
OS co-twin	23.23	13.34	11.28	634–636
SS co-twin	22.07	13.11	10.87	1390–1394
$\chi^2_{\text{diff}}$ (1 df)	17.07	1.79	3.92	
<i>p</i>	<.0001	.18	.05	
Effect size	.18	-.06	-.12	
Standard deviations				
Females	6.40	3.72	3.04	
Males	6.48	3.56	3.48	
Effect size, female vs. male SS twins	.66	-.48	-.53	

<sup>a</sup> OS = opposite-sex, SS = same-sex. Means shown are raw means over all individual twins in the category;  $\chi^2$ 's and *p*-values are from a model-fitting test of equality of OS and SS twins for that sex (age included in model); effect sizes are differences between means in standard deviation units; for co-twins, they are positive when the twins with OS co-twins score in the direction of the opposite-sex mean; for the sexes, a positive sign means that females score higher.

**Table V.** Younger Cohort: Correlations of Twins on Three Masculinity-Femininity Scales in Five Groups, Mean Ages, and Reliabilities of the Scales<sup>a</sup>

Group	Worried	Reserved	Breaks rules	Mean age	Ns
MZf	.36	.37	.42	23.3	454
MZm	.35	.30	.28	22.9	247
DZf	.13	.08	.28	23.2	310
DZm	.19	.06	.18	23.1	162
DZmf	.14	.15	.17	23.2	330
Reliabilities					
Alpha—f	.78	.65	.65		1062
Alpha—m	.77	.60	.64		635

<sup>a</sup> MZ = monozygotic, DZ = dizygotic, f = female, m = male. Pearson correlations between first and second twin. Ns for correlations are number of pairs, for reliabilities are number of individuals. Alphas are Cronbach alpha internal-consistency reliability coefficients, based on one twin from each pair.

to be more MZs and females than DZs and males, and the correlations for the MZ groups tend to run higher than those for the DZ groups. As before, the DZ groups tend not to differ very systematically among themselves, although one particular correlation, that

for DZ females on Breaks rules, is on the high side in both samples.

Note that within this age range there seem not to have been systematic age differences between the male and female twins participating in the study. In neither this nor the older cohort did age differ by zygosity. The reliabilities are a little lower in the younger than in the older cohort, except for rule-breaking, but remain roughly comparable for males and females.

Tables VI and VII show the same model-fittings as Tables II and III. (Again it was the case that an *aie* model fit better for Worried and Reserved, and an *ace* model for Breaks rules). The fit statistics showed negligible improvements in fit from allowing different genes to affect a trait in the two sexes (see first two rows of the table), so the simplifying hypothesis of the same genes could again be made. The overall chi squares are roughly comparable in the two cohorts for Worried and Reserved, but are higher in the younger cohort for Breaks rules. The other fit indices, however, suggest that all the model fits remain satisfactory, although the upper 90% confidence limit for Breaks rules slips into the “acceptable” rather than the “excellent” range for that index.

The sex differences in the genetic and environmental parameters are non-significant again for Reserved, but now are also so for Worried. Sex differences remain significant for Breaks rules; in both cases it is the females for whom the shared environment parameter is appreciable, reflecting their higher DZ correlation. However, the  $c^2$  estimate for females is fairly small ( $.30^2 = .09$ ).

The correlations with age tend to be lower in the younger cohort, which of course has a much smaller age range. The signs remain the same for Reserved and Breaks rules—the older twins tend to be more reserved and more conscientious. Again, consistent age trends are not present for Worried. The sex difference observed in the older cohort is reversed in the younger (the males now become more worried, the females less so, with age), but it is no longer statistically significant (Table VII). The sex difference for the relation of Breaks rules to age is also no longer present.

The corrected heritabilities are again in the neighborhood of .50, but tend to be slightly less than the corresponding ones from the older sample.

Table VIII shows chi-square tests of between-sample differences, obtained by constraining the specified parameters to be the same across samples in a joint model-fitting, while the others were left free to vary. The test in the first row of the table, for example, means that

**Table VI.** Younger Cohort: Fit Statistics, Parameter Estimates, and Heritabilities<sup>a</sup>

Parameter	Worried	Reserved	Breaks rules
Fit statistics			
$\chi^2$ (28df)	25.98	22.27	44.77
$\chi^2$ (29df)	26.00	22.27	45.07
RMSEA-ul90%	.038	.028	.064
SRMR	.074	.048	.054
Parameter estimates			
$a_f$	.50	.46	.58
$i_f$	.33	.39	.0
$c_f$	.0	.0	.30
$e_f$	.80	.79	.75
$a_m$	.58	.46	.53
$i_m$	.00	.24	.0
$c_m$	.0	.0	.04
$e_m$	.81	.84	.84
$r_{age,m}$	.08	.02	-.09
$r_{age,f}$	-.02	.12	-.07
Heritabilities			
$h^2_f$	.36	.37	.34
$h^2_m$	.34	.28	.28
$h^2_{f,corr}$	.46	.57	.52
$h^2_{m,corr}$	.44	.47	.44

<sup>a</sup> Parameters: a = additive genetic, i = epistatic genetic, c = shared environment, e = residual (unshared environment, error, interaction, etc.);  $r_{age}$  = fitted within-group correlation of trait with age. .0 = parameter fixed to zero. Subscripts: f = female, m = male, corr = corrected. RMSEA-ul90% = upper limit of 90% confidence interval for root mean squared error of approximation; SRMR = standardized root mean squared residual,  $h^2$  = heritability.

**Table VII.** Younger Cohort: Chi-Square Tests of Parameter Differences Between Sexes

Test	df	Worried	Reserved	Breaks rules
Same a, i/c, e	3	.40	1.97	12.15
<i>p</i>		.94	.58	.01
Same $r_{age}$	1	2.98	6.72	.06
<i>p</i>		.08	.01	.81

in the model fitting  $a_m$  was constrained to be the same in both samples, as was  $a_f$  (although they could be different from each other); likewise with  $e_m$  and  $e_f$  and  $i_m$  and  $i_f$  (or  $c_m$  and  $c_f$ ). Age regressions, means, etc., were left free to vary across samples. In the analyses in the later rows of the table, these in turn were constrained.

As is evident, the parameters or groups of parameters tested all showed significant differences across the two samples. Thus age or other differences between the samples did matter, although many of these differ-

**Table VIII.** Chi-Square Tests of Parameter Differences Between Cohorts

Test	df	Worried	Reserved	Breaks rules
Same a, i/c, e	6	17.66	14.01	28.13
<i>p</i>		.01	.03	.0001
Same $r_{age}$	2	11.70	11.35	7.87
<i>p</i>		.003	.003	.02
Same mean	2	37.02	16.09	75.65
<i>p</i>		<.0001	.0003	<.0001

ences are not large in absolute terms, as one can judge from inspection of the corresponding parameter values in the various tables.

Finally, and most importantly, how do the effects of having had a same-sex or opposite-sex co-twin compare in the younger and older cohorts? Table IX presents the same analysis as Table IV. First, it is no longer the case that for females the sex of the co-twin has no effect—women with opposite-sex co-twins score significantly higher on the rule-breaking scale, although the difference is small (effect size of .10). (The difference in the older cohort, though not statistically significant, was in the same direction.) Second, it is no longer true that males with opposite-sex co-twins are more feminine on the Worried dimension—indeed they are almost-but-not-quite-significantly *less* feminine. Results remain the same as before on the Reserved scale—sex of co-twin does not make a significant difference for either males or females. Lastly, it is again found that males with female co-twins score higher on the rule-breaking scale than do males with male co-twins, although the effect is still not very large (about one sixth of a standard deviation). Again the effects within the sexes are small relative to the between-sex differences.

**DISCUSSION**

The present negative results for two of the three masculinity-femininity dimensions are in line with those of Rose *et al.* (1994) on feminine attitudes and with the results of two smaller studies on sex-typed childhood play (Henderson and Berenbaum, 1997; Rodgers *et al.*, 1998). For the third scale, rule breaking, the results were borderline, at best. Thus it is clearly no longer appropriate (if it ever was) to claim general masculinization of the brain as an automatic sequel to a female having shared a uterus with a male. More limited aspects of masculinization remain a possibility. If

**Table IX.** Younger Cohort: Means on Three Masculinity-Femininity Measures for Twins with Opposite-Sex and Same-Sex Co-Twins<sup>a</sup>

Group	Worried	Reserved	Breaks rules	N
Females with				
OS co-twin	25.34	10.48	10.84	477–483
SS co-twin	26.05	10.48	10.51	1773–1782
$\chi^2_{\text{diff}} (1 \text{ df})$	3.32	.90	4.69	
<i>p</i>	.07	.34	.03	
Effect size	.12	.00	.10	
Males with				
OS co-twin	20.82	12.92	12.71	399–405
SS co-twin	21.73	12.45	12.13	1041–1055
$\chi^2_{\text{diff}} (1 \text{ df})$	3.81	1.85	11.27	
<i>p</i>	.05	.17	.0008	
Effect size	-.15	-.14	-.17	
Standard deviations				
Females	6.16	3.46	3.32	
Males	6.08	3.37	3.50	
Effect size, female vs. male SS twins	.71	-.58	-.48	

<sup>a</sup> OS = opposite-sex, SS = same-sex. Means shown are raw means over all individual twins in the category;  $\chi^2$ s and *p*-values are from a model-fitting test of equality of OS and SS twins for that sex (age included in model); effect sizes are differences between means in standard deviation units; for co-twins, they are positive when the twins with OS co-twins score in the direction of the opposite-sex mean; for the sexes, a positive sign means that females score higher.

masculinization of some brain functions does occur in response to prenatal exposure to a co-twin's hormones, a good deal of careful research will be required to elucidate when such effects do and do not happen.

More unusual is the finding of several differences on masculinity-femininity dimensions among *males*, as a function of the sex of their co-twin. It seems unlikely that the prenatal presence of a female co-twin has somehow simply attenuated the normal masculinization process in a male. For example, in the case of rule-breaking, where the difference was found in both samples, it was in the direction of males with female co-twins being *more*, not less, masculine. For the Worried scale, the effects, such as they were, lay in opposite directions in the two samples. For Reserved, there was no significant effect in either. The possibility of some subtle or localized prenatal effect can never be completely excluded, but, clearly, the explanation cannot be a simple one. Alternatively, one may consider postnatal explanations involving the social influence of a sister—although the question of why this should occur for some masculinity-femininity traits and not others arises here also.

In a number of instances, the results of model-fitting differed between the older and the younger twin samples. This is perhaps unsurprising, given that different aspects of masculinity and femininity appear to vary with age, and vary differently. It should also be noted that although we have referred to them as age differences, some of them might well be generational ones. The younger twin sample grew up in a society changed in many ways from the one in which the bulk of the older sample grew up, with different attitudes toward sex roles and traditional codes of morality.

It does not appear that we are dealing with fundamental differences in measurement across the cohorts or the sexes. The alpha reliabilities were all in reasonably good agreement, and were consistent with test-retest reliabilities. There was no evidence for genetic sex-limitation in either cohort, and the model fits were excellent in both, with the exception of rule-breaking in the younger cohort, where they were acceptable. The overall estimates of broad heritability were roughly similar across traits, across cohorts, and across the sexes (a little lower in the younger cohort, and for the males). Adjusted for error of measurement, the heritabilities appeared similar to those of around .50 typically reported for personality traits in twin studies.

There was no evidence of shared environmental effects,  $c^2$ , for the Worried or Reserved dimensions, nor for males on Breaks rules, although there was a positive estimate of  $c^2$  for females on the latter trait. Thus hormonal levels as such are plausible as a factor only for Breaks rules in females. The direction of the difference between females with same-sex and opposite-sex co-twins was consistent with such a hypothesis in both samples (those with male co-twins were higher) but it was small, and statistically significant only in the younger cohort. Moreover, one would need to have a different hypothesis for males, for whom  $c_2s$  were negligible, and for whom the presence of female co-twins led to higher, not lower, scores on this trait in both cohorts.

A positive finding for Breaks rules in females seems consistent with Resnick *et al.*'s (1993) finding of a positive effect for females on sensation-seeking, although our result for males would be puzzling here.

In addition to those differences already discussed, the sexes and cohorts showed some differences at the level of the detailed model parameters. Given that twin data have limited power to distinguish, for example, between additive and nonadditive genetic variation, and that differential recruitment by age and sex into

volunteer twin samples is likely, it would perhaps be prudent to await replication of these differences before speculating at length about their interpretation.

There are a number of reasons why negative results might be obtained in testing the hypothesis of prenatal exposure to hormones of the opposite sex. Several of these are discussed by Henderson and Berenbaum (1997).

First, it might be that hormonal effects are critical in accounting for normal male-female differences, but that they are threshold-dependent, and that in humans the diffusion of hormones across fetal membranes is too slight to reach such thresholds for most traits. (One might posit trait-to-trait variation in thresholds, or varying sensitivity during different periods of development, to account for the fact that some traits are affected and others not.)

Second, it may be that the particular traits under consideration do not differ sufficiently between the sexes to make it plausible that hormones affect them at all. This would not appear to be a likely explanation in the present study. The obtained male-female differences fell in the range of about a half to about two-thirds of a standard deviation; they were at least as large in the cross-validation sample as in the one in which the scales were originally derived. Some traits for which positive results have been reported show sex differences larger than this, but not dramatically so. For example, a measure of otoacoustic emissions showed an effect size of .71 between males and females, and one of .37 for heterosexual versus homosexual females (McFadden & Pasanen, 1998). A meta-analysis of the mental rotations test found an average effect size for males versus females of about .90 (Masters & Sanders, 1993). Resnick *et al.* (1993) reported effect sizes for gender ranging from .10 to .83 for the subscales of the Sensation-Seeking Scale. Moreover, many of the negative findings occur for traits with large between-sex differences. For play with sex-typed toys, Berenbaum and Snyder (1995) reported sex-difference effect sizes of 1.92 and 1.44 for boys' toys and girls' toys, respectively. Rodgers *et al.*'s corresponding effect sizes were .33 and .58. Rose *et al.* (1994) had sex differences of more than 3 standard deviations for feminine interests. And Loehlin's and Martin's (1998) study of reproduction-related variables dealt with traits with the ultimate sex difference in magnitude: they do not occur in males at all.

Finally, of course, it may be that hormonal effects differentially affecting same-sex and opposite-sex pairs do not occur in humans, or occur, but at such low levels

as to make them impossible to establish convincingly with less than astronomical sample sizes. On this view, the scattered positive results appearing in the literature may reflect postnatal socialization effects, chance results appearing in small samples, or developmentally very special cases. Further research may support or refute or clarify such an interpretation.

In the meantime, the results of the present study suggest continued interest in the possible effects of prenatal hormones on rule-breaking tendencies in females. They strongly suggest that in future studies of opposite-sex and same-sex co-twins males should be included. And they suggest that age is an important parameter to be taken into account in any such research.

## ACKNOWLEDGMENTS

We are grateful to John Pearson for assistance with the data, and to S. A. Berenbaum and J. M. Bailey for helpful comments on an earlier version of this paper.

## REFERENCES

- Berenbaum, S. A., and Snyder, E. (1993). Early hormonal influences on childhood sex-typed activity and playmate preferences: Implications for the development of sexual orientation. *Devel. Psychol.* **31**:31-42.
- Browne, M. W., and Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K. A., and Long, J. S. (eds.), *Testing Structural Equation Models*, Sage, Thousand Oaks, CA, (pp. 136-162).
- Burns, E. M., Arehart, K. H., and Campbell, S. L. (1992). Prevalence of spontaneous otoacoustic emissions in neonates. *J. Acoust. Soc. Am.* **91**:1571-1575.
- Cloninger, C. R., Przybeck, T. R., and Svrakic, D. M. (1991). The Tridimensional Personality Questionnaire: U.S. normative data. *Psychol. Rep.* **69**:1047-1057.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. Ed.), Academic Press, New York.
- Cole-Harding, S., Morstad, A. L., and Wilson, J. R. (1988). Spatial ability in members of opposite-sex twin pairs. *Behav. Genet.* **18**:710 (abst.).
- Collaer, M. L., and Hines, M. (1995). Human behavioral sex differences: A role for gonadal hormones during early development? *Psychol. Bull.* **118**:55-107.
- Eysenck, S. B. G., Eysenck, H. J., and Barrett, P. (1985). A revised version of the Psychoticism scale. *Person. Individ. Diff.* **6**:21-29.
- Henderson, B. A., and Berenbaum, S. A. (1997). Sex-typed play in opposite-sex twins. *Devel. Psychobiol.* **31**:115-123.
- Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equat. Mod.* **6**:1-55.
- Jardine, R., Martin, N. G., and Henderson, A. S. (1984). Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genet. Epidemiol.* **1**:89-107.
- Lippa, R. (1995). Do sex differences define gender-related individual differences within the sexes? Evidence from three studies. *Pers. Soc. Psychol. Bull.* **21**:349-355.

- Loehlin, J. C., and Martin, N. G. (1998). A comparison of adult female twins from opposite-sex and same-sex pairs on variables related to reproduction. *Behav. Genet.* **28**:21–27.
- Loehlin, J. C., Spurdle, A., Treloar, S. A., and Martin, N. G. (in press). Number of X-linked androgen receptor gene CAG repeats and femininity in women. *Person. Individ. Diff.*
- Lykken, D. T., Tellegen, A., and DeRubeis, R. (1978). Volunteer bias in twin research: The rule of two-thirds. *Soc. Biol.* **25**:1–9.
- Masters, M. S., and Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behav. Genet.* **23**:337–341.
- McFadden, D. (1993). A masculinizing effect on the auditory systems of human females having male co-twins. *Proc. Natl. Acad. Sci. USA* **90**:11900–11904.
- McFadden, D., and Pasanen, E. G. (1998). Comparison of the auditory systems of heterosexuals and homosexuals: Click-evoked otoacoustic emissions. *Proc. Natl. Acad. Sci. USA* **95**:2709–2713.
- Miller, E. M. (1994). Prenatal sex hormone transfer: A reason to study opposite-sex twins. *Person. Individ. Diff.* **17**:511–529.
- Neale, M. C., and Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families*. Kluwer, Dordrecht.
- Resnick, S. M., Gottesman, I. I., and McGue, M. (1993). Sensation seeking in opposite-sex twins: An effect of prenatal hormones? *Behav. Genet.* **23**:323–329.
- Rodgers, C. S., Fagot, B. I., and Winebarger, A. (1998). Gender-typed toy play in dizygotic twin pairs: A test of hormone transfer theory. *Sex Roles* **39**:173–184.
- Rose, R. J., Kaprio, J., Winter, T., and Koskenvuo, M. (1994). Femininity and fertility in sisters of twin brothers: Neither prenatal androgenization nor cross-sex socialization. Paper presented at meetings of the International Society for the Study of Behavioral Development, Amsterdam, 28 June–12 July.
- Steiger, J. H. (1989). *EzPATH: Causal Modeling*. SYSTAT Inc., Evanston, IL.
- vom Saal, F. S. (1989). Sexual differentiation in litter-bearing mammals: Influence of sex of adjacent fetuses in utero. *J. Anim. Sci.* **67**:1824–1840.