This paper appeared in the Journal of Child Psychology and Psychiatry (2005)

Gender-typed Play and Amniotic Testosterone

Abstract

Sex differences in play are apparent in a number of mammalian species, including humans. Prenatal testosterone may contribute to these differences. We report the first attempt to correlate gender-typed play in a normative sample of humans with measurements of amniotic testosterone (aT). Testosterone was measured in the amniotic fluid of 53 children (31 male, 22 female). A strong sex-difference was observed in aT and, at age 4.75 to 5.8 years, on a modified version of the Child Game Participation Questionnaire (CGPQ). Hierarchical regression analyses on the entire group and within sex correlations suggested that variations in aT did not contribute to individual differences in game participation, as reported by the mother. A critique of explanations for this finding is presented.

Gender-typed Play and Amniotic Testosterone

Sex differences in play patterns are found in many species, including human beings. Boys engage in more rough play, athletic games and prefer construction and transportation toys, while girls show more play-parenting and prefer play with dolls and kitchen supplies (DiPietro, 1981; Humphreys & Smith, 1984; Pellegrini & Smith, 1998). Exposure to hormones in the prenatal period may organize future sex differences that then require no circulating hormones for their expression (Phoenix, Goy, Gerall, & Young, 1959). Such organizational effects may predispose individuals to behave in a certain way but do not act independently of social and contextual influences.

The occurrence of rough and tumble play is related to prenatal or early postnatal exposure to male hormones in many mammalian species (Pellis, 2002; Young, Goy, & Phoenix, 1964). Testosterone (T) implants into the amygdala during the neonatal period masculinize play in juvenile female rats (Meaney & McEwen, 1986). Conversely, exposure to antiandrogens such as vinclozolin and flutamide either during prenatal or early neonatal life, feminizes play behavior in male rats (Casto, Ward, & Bartke, 2003; Hotchkiss, Ostby, Vandenbergh, & Gray, 2003). As reviewed by Wallen (1996), female rhesus monkeys exposed to a long period of prenatal androgen or to a short period of androgen late in pregnancy show a dramatic increase in the amount of rough and tumble play. Exposing males to higher prenatal androgen levels had no effect (perhaps because their natural levels were already high). Suppressing testicular function after birth did not affect rough and tumble play in males, suggesting that this preference had its root in the prenatal period.

Gender-typical play has been examined in girls with elevated prenatal androgen exposure as a result of classic congenital adrenal hyperplasia (CAH) (non-classical forms will not be discussed in this article). Most girls with CAH are recognized at birth or during early childhood at which point the hormonal abnormalities can be ameliorated through cortisone-replacement therapy (Hines, 2002; New, 2003; White & Speiser, 2002). As a result, for girls who received early neonatal diagnosis and consistent hormone treatment, androgen is elevated only in the prenatal (at least from week 10) and early neonatal period. Studies indicate that T exposure is in the normal male range during midgestation (Carson et al., 1982; Forest, Betuel, Couillin, & Boue, 1981). Girls with CAH show increased play with boy's toys and decreased play with girl's toys when compared to matched controls or their unexposed female relatives (Berenbaum & Hines, 1992; Berenbaum & Snyder, 1995; Dittman et al., 1990; Nordenstrom, Servin, Bohlin, Larsson, & Wedell, 2002; Servin, Nordenstrom, Larsson, & Bohlin, 2003; Zucker et al., 1996). They also show less interest in infants (Leveroni & Berenbaum, 1998).

Boys with CAH appear to have normal levels of androgens prenatally (Pang et al., 1980; Pang, Levine, Chow, Faiman, & New, 1979) and do not differ from their male relatives in play with boy's and girl's toys (Berenbaum & Hines, 1992; Berenbaum & Snyder, 1995). The levels of circulating prenatal T in males with CAH are complex, making it substantially more difficult to test specific hypotheses about testosterone-behavior relations in this group.

Studies of girls with CAH provide some of the most compelling evidence that prenatal androgens have a lasting effect on human behavior, including gender-typical play. However, interpretation of the results could be confounded by difficulty in differentiating between the effects of elevated prenatal androgens and other characteristics of the condition. Girls with CAH are also exposed to high levels of adrenocorticotropic hormone (ACTH) and treatment may result in glucocorticoid excess (Berenbaum, 2001). In contrast to the androgen hypothesis, there is no clear theoretical reason why changes in ACTH would masculinize play behavior. Masculinization appears to be correlated with degree of androgen excess (Nordenstrom et al., 2002; Servin et al., 2003).

There is also evidence suggesting that foetal testosterone levels affect play behavior in boys. Medroxyprogesterone acetate (MPA) is a synthetic progestin that decreases T levels. Rough and tumble play is reduced in male rats exposed neonatally to MPA (Birke & Sadler, 1983). Boys exposed to MPA for at least 1 week during the second to eighth month of pregnancy showed some demasculinization and feminization of play on a scale derived from the CGPQ (Meyer-Bahlburg, Feldman, Cohen, & Ehrhardt, 1988). Exposure to polychlorinated biphenyls (PCBs) also suppresses T levels (Hany et al., 1999; Kaya et al., 2002). In a study of Dutch school children, higher prenatal exposure to PCBs was related to less masculine play as assessed by the Pre-School Activity Inventory (PSAI) (Vreugdenhil, Slijper, Mulder, & Weisglas-Kuperus, 2002).

In the present paper we report the first study to directly measure amniotic testosterone (aT) and relate it to play behavior in a normative sample of children. The majority of studies showing that variations in fetal T are related to differences in gender-typed play have used groups with large abnormalities in prenatal endocrine conditions due to genetic flaws or fetal exposure to exogenous substances. Results of these studies may not be generalizable to populations whose prenatal T exposure is within normal physiological limits. Even when this is not an issue, variations in T levels are inferred but not measured quantitatively.

T can be measured in amniotic fluid collected during mid-trimester amniocentesis (Finegan, Bartleman, & Wong, 1989). T is thought to enter the amniotic fluid via diffusion through the skin in early pregnancy, and later from fetal urination (Klopper, 1970; Nagamani, McDonough, Ellegood, & Mahesh, 1979; Robinson, Judd, Young, Jones, & Yen, 1977). Although the exact correlation between T levels in the serum and the amniotic fluid is unknown, the maximal sex difference in aT occurs between weeks 12 and 18, closely paralleling peak serum levels (Finegan et al., 1989). Amniocentesis is almost exclusively performed in the second trimester, around week 16 of gestation. If sexual differentiation of the human brain did not occur until late in pregnancy, aT levels would be unlikely to relate to later sex-typical behavior. Therefore, it is important to consider the timing of sexual differentiation in the human brain.

There are substantial inter-species differences in the timing and process of sexual differentiation of steroid-sensitive regions of the brain and behavior (Wallen & Baum, 2002). In animal models, the general critical period for steroid mediated sexual differentiation of the brain usually occurs when sex-differences in serum T are highest (Smith & Hines, 2000). Therefore it is likely that this is an important period for sexual differentiation of the brain in humans as well. Udry, Morris, & Kovenock (1995) report that androgen exposure (measured in maternal blood) in the second (and no other) trimester of fetal life, in interaction with adult androgens, masculinized women’s behavior. However, the second trimester is not necessarily the only period during which differentiation occurs. T production does decline in males by the third trimester of pregnancy (Abramovich & Rowe, 1973; Reyes, Boroditsky, Winter, & Faiman, 1974) and reported sex differences in T are minimal or absent near term (Beck-Peccoz et al., 1991; Carson et al., 1982; Nagamani et al., 1979; Warne, Faiman, Reyes, & Winter, 1977), but sex differences in T levels measured in serum or umbilical cord blood (Forest & Cathiard, 1975; Forest, Sizonenko, Cathiard, & Bertrand, 1974) have been reported at birth. There is another T surge in early neonatal life and again at puberty (MacLusky & Naftolin, 1981).

The children in this study are part of a longitudinal project. Earlier studies have focused on social-communicative skills. aT was negatively related to vocabulary size at 12 months of age (Lutchmaya, Baron-Cohen, & Raggatt, 2002a), amount of eye contact at 12 months of age (Lutchmaya, Baron-Cohen, & Raggatt, 2002b), and quality of social relationships at age 4 (Knickmeyer, Baron-Cohen, & Raggatt, in press). An independent group has shown that at age 8 girls with higher levels of aT performed a mental rotation task faster than girls with lower levels (Grimshaw, Sitarenios, & Finegan, 1995). At age 10, girls with higher levels of aT showed a more masculine pattern of cerebral lateralization (Grimshaw, Bryden, & Finegan, 1995). No prior studies have examined sex typical play in relation to aT. Thus, our data extends previously studied relationships between aT levels and human behavior to include sex-typical play.

We measured gender-typed play via maternal report with a questionnaire adapted from the Child Game Participation Questionnaire (CGPQ). We predicted that boys and girls would show strong differences in their scores for male-typical and female-typical games. We also predicted, based on previous research in humans, that aT would be positively correlated with scores for male-typical games and negatively correlated with scores for female-typical games.

Methods

Participants

Participants were 53 children (31 male, 22 female), age 4.75 to 5.8 years at the time of behavioral assessment. The final sample of 53 subjects represent those who responded from a larger sample of 80 families taking part in a long-term study on the effects of aT (the response rate for the initial recruitment was approximately 50%). Mothers had undergone amniocentesis in the Cambridge region between June 1996 and June 1997 and had given birth to healthy singleton infants between December 1996 and December 1997. The majority of mothers were referred for amniocentesis because of late maternal age (25%) or high results on the triple test (indicating an increased risk for Down's syndrome) (60%). All amniotic samples tested negative for Down's and other chromosomal abnormalities. Children who have had amniocentesis show no evidence of decreased well being or impaired brain development (Finegan, Sitarenios, Bolan, & Sarabura, 1996). Any child whose medical records indicated ill health at birth, for example requiring long stays in the SCBU (Special Care Baby Unit), were excluded from the study. All subjects were European with the exception of one male, who was Asian (Indian, Pakistani, or Bangladeshi). The mean maternal age for the entire sample was 35.0 years (sd = 4.85). Maternal education level was measured according to a 5 point scale: 1 = no formal qualifications, 2 = 'O' level / G.C.S.E. (A General Certificate of Secondary Education represents successful completion of exams after 16 years schooling) or equivalent, 3 = 'A' level, HND, or other vocational qualification (A levels are generally in academic subjects while the Higher National Diploma is awarded for vocational subjects. Both usually require 18 years schooling + exams), 4 = university degree, 5 = postgraduate qualification. Mean maternal education for the entire sample was 3.25 (sd = 0.87).

Outcome Variable

Mothers completed a modified version of the Child Game Participation Questionnaire (CGPQ) (Bates & Bentler, 1973) (See appendix A for pilot study results). The questionnaire included 10 masculine items, 10 feminine items, and 8 neutral items. For each game, mothers indicated their child's interest on a Likert scale (1 = not at all interested to 5 = very interested). A femininity score was calculated by adding together a child's scores over all feminine items (1=0, 2=1, 3=2, 4=3, 5=4). A masculinity score was calculated by adding together a child's score over all masculine items in the same way. The femininity and masculinity scores had a possible range of 0 to 40.

Predictor variables

Amniotic testosterone levels (aT) (nmol/L). The predictor of greatest interest in this study is amniotic testosterone. T levels in amniotic fluid were measured by radioimmunoassay by the Department of Clinical Biochemistry, Addenbrooke's Hospital Cambridge, a method our group has reported previously (Knickmeyer et al., in press; Lutchmaya et al., 2002a; Lutchmaya et al., 2002b)(see appendix B for details of the assay). There were significant differences between boys' and girls' T levels, t (43) = 7.1, p = 0.00, d = 1.8. Equal variances were not assumed on any t-tests. The probability of a type I error was maintained at 0.05 for all t-tests. If the lowest aT levels in our sample were near the detection limit of the assay (0.1 nmol/L), it would raise the possibility of a floor effect (particularly for girls). We further investigated the distribution of scores to determine whether this was the case. No girls had undetectable levels of aT. Only 2 girls (about 9% of the female sample) scored below 0.2 nmol/l, indicating that there was not a strong floor effect. However, the distribution of female scores was skewed to the left in comparison to the distribution of male scores. Transformation of scores was not necessary. There was also a degree of overlap between fT levels in boys and girls in this study, which raises the possibility that fT levels were declining in males. The mean fT levels in both males and females (1.02 and 0.39 nmol/l respectively) were slightly lower in our study than those in Finegan et al. (1989) (1.34 and 0.58 nmol/l respectively). The effect size for the sex-difference in fT was also lower in our study, $d = 1.87$ vs. $d = 2.7$.

We also included the following control variables in our analysis.

Prenatal estrogen levels (pmol/L). Estradiol is the most biologically active endogenous estrogen. In rodents it masculinizes and defeminizes the brain when it is synthesized in vivo via aromatization of T and related precursors, although in some cases T directly masculinizes the brain (see De Vries & Simerly, 2002 for review). Studies of individuals with complete androgen insensitivity syndrome and of girls exposed in utero to the synthetic estrogen, diethylstilbestrol (DES), suggest that in humans, T directly influences sexual differentiation without being converted to estrogen (Hines, 2002; Hines, Ahmed, & Hughes, 2003). Amniotic estradiol levels were also assayed by the Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge (Appendix B). There were no significant differences between estrogen levels in boys and girls, $t(48) = 0.31$, $p = 0.76$, $d = 0.08$. Prenatal estrogen level was positively skewed (skewness = 2.05). A natural logarithmic transformation was carried out, which reduced the skewness considerably (skewness = 0.93). The transformed variable was used in all subsequent correlations and regressions. There were no significant differences between males and females when the transformed version of the variable was used, t (36) = 0.14, p = 0.89, d = 0.00.

Prenatal alpha-fetoprotein level (MU/L). Alpha-fetoprotein (AFP) is thought to be a general marker for severe fetal ill health and also provides a specific control for any unexpected abnormalities of amniotic fluid dilution (Wathen, Campbell, Kitau, & Chard, 1993). Amniotic AFP levels were also assayed by the Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge (Appendix B). There were no significant differences between AFP levels in boys and girls, t (51) = 0.09, p = 0.93, d = 0.02.

Sex of Child. Boys were coded as 1 and girls were coded as 2 for all analyses.

Gestational age at amniocentesis (weeks). Levels of aT vary during gestation. Although amniocentesis occurs on average at week 16, it can occur as early as week 12 and as late as week 22. Therefore it was important to determine whether aT was related to gestational age in our sample. Gestational age at amniocentesis (as calculated from date of last menstrual period) was obtained from hospital records. Males showed no significant linear relationship between aT and gestational age, r (24) = 0.21, p = 0.33. No quadratic relationship was apparent. For females, there was a significant linear relationship, r (21) = -0.59 p = 0.01. This was an unexpected finding, given that Reyes et al. (1974) report no change in fetal serum concentrations of T for females during this same period. One girl had a fT level 2 standard deviations above the mean and a gestational age 2 standard deviations below the mean. When this case is removed the correlation with gestational age is no longer significant, *r* (20) = -0.42 *p* = 0.07. This case was excluded from the regression analyses.

Sociodemographic variables. A range of sociodemographic variables were also included in this study because of their possible importance in determining the child's social environment, these included maternal age, maternal education, and gender of older siblings.

Both maternal age and maternal education have been found to influence a person's own attitudes toward gender roles; and this could affect the degree to which mothers encourage and expect children to show gender-typical play. Older persons tend to be more traditional in their gender-role perceptions. The higher the level of education, the more liberal people are in their gender-role perceptions (Corder & Stephan, 1984; Kulik, 2002; Lackey, 1989; Quarm, 1983; Togeby, 1995). There were no significant sex differences in maternal age between boys and girls, t (45) = 1.13, p = 0.27, d = 0.28. There was also no significant difference in maternal education between boys and girls, t (23) = 0.12, p = 0.90, d = 0.06.

Gender of older siblings was examined because several studies show that sex-typed behavior may be influenced by the sex of a sibling, especially an older one (Abramovitch, Corter, Pepler, & Stanhope, 1986; Henderson & Berenbaum, 1997; Stoneman, Brody, & MacKinnon, 1986). For the variable "older brothers," children were coded 1 if they had older brothers (regardless of number) and 0 if they had no older brothers. For the variable "older sisters," children were coded 1 if they had older sisters (regardless of number) and 0 if they had no older sisters. Chi-square was used to test whether girls and boys differed in the number of children having older brothers and in the number of children having older sisters and this analysis was not statistically significant, $\chi^2(1, N = 43) = 0.69$, $p = 0.41$ and $\chi^2(1, N = 43) = 0.37$, $p = 0.54$ respectively. Data on gender of older sibling was missing for approximately 30% of the girls and 10% of the boys. There was no obvious reason why data on this variable was missing for more girls than boys and this was probably due to chance. Chi-square was used to test whether significantly more girls were missing data on older siblings than boys. This analysis showed a trend, but was not statistically significant, $\chi^2(1, N = 53) = 2.80$, $p = 0.09$.

Results

Descriptive Statistics

Table 1 presents the mean, standard deviation, range, and gender effect size for predictor and outcome variables of each sex separately. Males scored higher on the masculinity scale t (41) = 12.7, p = 0.00, d = 3.6. Females scored higher on the femininity scale t (34) = -11.1, p = 0.00, d = 3.2. These scales were explored further to investigate whether variations in aT largely account for the observed sex-differences and individual variation within sex.

Hierarchical Regression Analyses

The first analyses explored the contribution of aT to scores on the masculinity and femininity scales. In block 1, sex was entered. In block 2, aT was entered. In block 3, the interaction of aT and sex was entered. Table 2 summarizes the results of these analyses. For the masculinity scale, inclusion of sex in the first stage produced a significant F change, F change = 161, $\beta = -0.87$, $p = 0.00$. This model explained 76% of the variance in masculinity scores. Inclusion of aT in the second stage did not produce a significant F change, F change = 0.06, $\beta = -0.02$, $p = 0.80$). Examination of the beta weight sizes suggests that aT does not account for a significant proportion of variance not accounted for by sex. The inclusion of a sex by aT interaction in the third stage also did not produce a significant F change, F change = 0.81, β = 0.24, p = 0.37. Examination of the beta weight sizes suggests that the interaction of aT and sex does not account for a significant proportion of variance not accounted for by sex. Residual analysis showed acceptable plots and no outliers. The only significant predictor in the final model was sex. However, to further investigate, we analyzed the relationship between the masculinity scores and aT within each sex. It should be kept in mind that this reduced the sample-size by half and therefore the power of the analysis. The masculinity scale did not correlate with aT levels in either boys or girls, r (31) = 0.00, p = 0.98; r (21) = -0.19, p = 0.40, respectively.

For the femininity scale, inclusion of sex in the first stage produced a significant F change, F change = 134, β = 0.85, p = 0.00. This model explains 72% of the variance in femininity scores. Inclusion of aT in the second stage did not produce a significant F-change, F change = 0.26, β = 0.05, p = 0.61. Examination of the beta weight sizes suggests that aT does not account for a significant proportion of variance not accounted for by sex. The inclusion of a sex by aT interaction in the third stage also did not produce a significant F change, F change = 3.91, β = -0.57, p = 0.054. The only significant predictor in the final model was sex. However, in the third stage, both aT and the sex by aT interaction approached significance, p = 0.058 and p = 0.054 respectively. Residual analysis showed acceptable plots and no outliers. We explored the potential interaction by analyzing the relationship between femininity scores and aT within each sex. It should be kept in mind that this reduced the sample-size by half and therefore the power of the analysis. The femininity scale did not correlate with aT levels in boys or girls, r (31) = -0.01, $p = 0.94$; r (21) = 0.36, $p = 0.10$, respectively.

Finally, we performed a hierarchical regression analysis which took into account the background variables we had measured. In block 1, any predictor variable that correlated significantly with the outcome variable at $p < 0.20$ was entered into the model (as recommended by Altman (1991)). Suppressor variables were also included when possible; these were predictors that correlated highly, $p < 0.01$, with the other predictors in the model, but were not significantly correlated with the outcome variable (See Table 3). In block 2, sex and aT were tested for inclusion using a step-wise analysis (Entry criteria was $p < 0.05$; removal criteria was $p > 0.1$). In block 3, the interaction of sex and aT was tested for inclusion using a stepwise analysis (entry and removal criteria as above). Table 4 summarizes the results of these analyses. The only significant predictor in both final models was sex. Residual analysis showed acceptable plots and no outliers.

Discussion

This study confirmed very large sex differences for both the masculinity and femininity scales of the Children's Play questionnaire. For comparison, in Meyer-Bahlburg, Sandberg, Dolezal, & Yager's (1994) factor analytic study of the CGPQ in 6-10 year old children, the masculinity scale had a d of 1.90 and the femininity/preschool scale had a d of 1.67. The majority of psychological studies demonstrate moderate effect sizes (i.e. $d = 0.5$) (Eagly, 1995).

We also predicted that T levels measured in amniotic fluid would be related to scores for male and female items. However, sex was the only significant predictor in the final models of our regression analyses for both male and female items. Within sex aT was not significantly correlated with scores on either scale, and *r* values were low. Our results suggest that prenatal testosterone, measured at this stage of development, is not related to individual variation in gender-typical play behavior. There are several possible explanations for this outcome, which reveal important factors to consider when investigating hormone-behavior relations.

(1) *T present at the time our samples were taken is related to gender-typical toy preferences, but amniotic testosterone may not be a reliable proxy measure for prenatal testosterone exposure of the brain.* Amniotic fluid studies of T make the assumption that amniotic levels are correlated with actual exposure levels, but there is no direct evidence to either support or contradict this assumption. It is important to note that in all existing studies, including the present one, hormones are assayed at a single timepoint. However, given that previous studies with this group have shown a relationship between aT and sex-dimorphic variables such as frequency of eye contact (Lutchmaya et al., 2002b), vocabulary development (Lutchmaya et al., 2002a), and quality of social relationships (Knickmeyer et al., in press), it seems likely that amniotic testosterone is an adequate proxy measure for actual exposure. These studies used similar analytical strategies to the present study, and all showed significant sex differences, $d = 0.53$, $d = 0.66$, and $d = 0.47$ for vocabulary size, frequency of eye contact, and quality of social relationships respectively.

(2) *T present at this period is related to gender-typical play, but the questionnaire we used did not accurately measure the children's behavior*. Relying on parental report has some drawbacks, including the possibility that different parents may interpret items differently. It is also possible that parents' reports reflected their

expectations of their child's behavior rather than the child's actual behavior. Lytton & Romney (1991), in a meta-analytic study of differential socialization of boys and girls, found that parents in North American studies encouraged sex-typed activities, thus lending credence to the idea that parents would assume their child had sex-typical play. However, a study by Meyer-Bahlburg, Erhardt, & Feldman (1985) showed that parental ratings were very similar to those produced when children themselves were given the CPGQ. Furthermore, observational studies of girls with CAH indicate that parents do accurately represent their children's gender atypical behavior (Nordenstrom et al., 2002). The problem of parental confounds could be overcome by a direct laboratory observation of toy play of the study group.

(3) ***T present at this period does contribute to gender-typical toy preferences, but the effect is only detectable when the child is exposed to highly atypical levels of prenatal T.*** It is possible that prenatal T contributes to the sex difference in toy preference, but is not easily detectable with the less extreme inter-individual variations that normally occur within sex. For example, an effect of prenatal T on play behavior is easily observed when a female is exposed to the abnormally high levels of prenatal T that occur in CAH or when a male is subject to abnormally low prenatal T. While the degree of masculinization and defeminization in girls with CAH is associated with severity of the disorder (Servin et al., 2003), we do not know how prenatal T variations occurring between CAH girls with different severities compares to the physiological variations that normally occur within either sex. Dosage effects could explain why studies of females with CAH show an effect on toy preference, but our amniotic fluid study did not. Grimshaw, Sitarenios, et al. (1995), in their study of aT and mental rotation, reported no

relationship between aT and spatial play. Several studies of play in opposite-sex twins have also yielded paradoxical results. The rationale for opposite-sex twin studies comes from experiments in rats: female rats adjacent to male rats in utero are masculinized (Clemens, 1974), possibly by T diffusing across the amniotic membrane (Fels & Bosch, 1971) or being carried through the maternal circulation. (Meisel & Ward, 1981). Human twin pregnancies are very different than multiple offspring pregnancies in rats, but T may transfer from the male to the female fetus through amniotic diffusion also in humans (Resnick, Gottesman, & McGue, 1993). The fetal skin is permeable to fluid and some dissolved solutes up to week 18 of gestation, and amniotic fluid moves through the entire fetoplacental unit (Abramovitch & Page, 1972; Brace & Resnik, 1989; Findlay, 1984). Human females with male co-twins have been reported to be masculinized with regards to sensation seeking (Resnick et al., 1993) and spatial ability on a mental rotation task (Cole-Harding, Morstad, & Wilson, 1988). However, studies of play preferences in opposite sex twins show no effect (Henderson & Berenbaum, 1997; Rodgers, Fagot, & Winebarger, 1998). Perhaps the level of T passed from male twins to female co-twins is not sufficient to produce changes in play.

(4) *The strength of the relationship between prenatal T and outcome may vary for different behaviors.* Differences in the effect size of prenatal T have important implications for sample size. If the effect of normal variation of prenatal T on play is very small, then only very large studies will reveal it. Most amniocentesis studies are based on relatively low sample sizes (55 in this study, 60 in Grimshaw, Sitarenios, et al. (1995)), but have consistently found significant relationships. Lutchmaya et al. (2002a) examined vocabulary size at 18 months in 87 children (40 girls and 47 boys). aT was a

significant predictor with a β of 0.6; sex was also a significant predictor with a β of 1.3 (using a backward stepwise regression analysis). In contrast, in the current study beta weight sizes for aT were much smaller. Grimshaw, Sitarenios, et al. (1995) found significant within-sex correlations for aT and mental rotation speed for both girls and boys despite small sample sizes, *r* (12) = 0.67 and *r* (13) = -0.62. Our within sex correlation analyses had sufficient power (greater than 0.80) to detect a similar effect size to that seen in Grimshaw, Sitarenios, et al. (1995), but would have had reduced power in detecting smaller effect sizes. Our multiple regression analysis had sufficient power (at least 0.80) to detect medium and large effect sizes (as defined by Cohen (1988)). A sample size of 250-500 would be needed to detect small effect sizes. Therefore, we cannot rule out a small effect of aT. However, the extremely low β and *r* values seen in the current study, suggest that even in a larger sample no stronger relationship between aT and play would be observed.

(5) *It is possible that prenatal T contributes to gender-typical toy preferences, but does so at a different time period than that examined in this study*. Various sex dimorphic behaviors have different organizational periods, i.e. the restricted temporal period of development during which brain tissues that mediate a given behavior can be modified. It is possible that the sensitive period for gender-typical play occurs later in development than the time our amniotic fluid samples were taken. This would fit in with the observation that female rhesus monkeys show a dramatic increase in rough play when exposed to a long period of prenatal androgen (≥35 days) or to a short period of androgen late in pregnancy (days 115-139), but not to a short period of androgen early in pregnancy (days 40-64) (Goy, 1978; Goy, 1981; Goy et al., 1988). Children with CAH

would be exposed to high levels of T throughout pregnancy, and thus would be exposed regardless of when the critical prenatal period for gender-typical play occurs in humans. Although midgestation has been considered the most important period in human sexual differentiation (Abramovich & Rowe, 1973; Finegan et al., 1989; Udry et al., 1995), sex differences in T levels have been reported at birth (Forest & Cathiard, 1975; Forest et al., 1974). There is also a surge in T in males during months 1-5 of the neonatal period (Chemes, 2001; Forest et al., 1974). The behavioral effects of this surge are currently unknown. It is possible that gender-typical toy preference is related to neonatal T levels as opposed to prenatal levels. In Berenbaum & Snyder's (1995) study of play in children with CAH, the median age at diagnosis for girls was 7 days, well before the neonatal surge. However, age at diagnosis ranged from 0 days to 5.3 years. A more recent study (Nordenstrom et al., 2002) separated girls with extremely late diagnoses (3-6 years) from girls diagnosed during the neonatal period and examined these separately, although they did not specify when they considered the neonatal period to end. This study supported both prenatal and postnatal effects of androgens. Berenbaum, Duck, & Bryk (2000) also examined prenatal versus postnatal androgen excess and found that sex-atypical play was associated with inferred prenatal exposure, but not with early inferred postnatal exposure, but note that androgen levels were not directly measured. If the timing of the critical period in rhesus monkeys is similar to that in humans, we can use those studies (Goy, 1978; Goy, 1981; Goy et al., 1988) and studies of CAH which compare prenatal and postnatal exposure (Nordenstrom et al. 2002; Berenbaum, Duck, & Bryk 2000), to narrow down the critical period for testosterone effects on gender-typed play to later pregnancy. Given that reported sex differences in T are minimal or absent near term

(Beck-Peccoz et al., 1991; Carson et al., 1982; Nagamani et al., 1979; Warne, Faiman, Reyes, & Winter, 1977), we would suggest a critical period in the late midtrimester or early third trimester.

(6) *Prenatal T does not contribute to the development of gender-typical toy preferences*. The changes in play behavior observed in CAH may be the result of other characteristics of the condition or because parents treat CAH daughters virilized at birth differently (Quadagno et al., 1977). There are clearly many potential factors, both biological and social that could produce sex differences in play. As the Lytton & Romney (1991) study shows, parents encourage their children to use sex appropriate toys. Children recognize "appropriate" toys and roles at an early age and emulate the behavior of same-sex models in preference to opposite-sex ones (Greif, 1976). As discussed earlier, there is currently no empirical support for any one factor, or the way they potentially interact (Berenbaum, 2001; Berenbaum & Hines, 1992; Dittman et al., 1990; Goy et al., 1988; Nordenstrom et al., 2002; Henderson & Berenbaum, 1997). Reviews of other variables which may explain gender-typed behavior can be found in (Hughes, 1991; Lippa & Hershberger, 1999; Powlishta, Sen, Serbin, Poulin-Dubois, & Eichstedt, 2001). These include peer influences, societal/media influences, and other biological factors.

Our study is also limited by the problems inherent in studying fetal endocrinology. We have assumed that aT levels accurately represent serum levels and brain exposure, but as discussed previously this assumption has not been tested empirically. In the serum, binding proteins and degradation enzymes affect the availability of the hormone. Only unbound T is biologically active. The assay used in this study measured total T in the amniotic fluid. However, because T is thought to enter the amniotic fluid via fetal urination and bound T is protected from excretion in the urine, the amniotic levels should primarily reflect unbound T. Presence and sensitivity of appropriate receptors also determines whether and how potent T’s effects may be.

Determining gestational age at amniocentesis is not exact. We calculated gestational age using the date of last menstrual period. More accurate estimates might be obtained using sonographic measures (such as femur length), but these are available for fewer children.

A final limitation of research using this method is that a truly random sample cannot be collected, since one can only include individuals who have decided/been advised to have an amniocentesis due to late maternal age or other factors that increase the risk of fetal abnormality. Previous studies investigating the relationship of prenatal T to cognitive development in humans have relied upon individuals with abnormal hormonal environments during pregnancy or those exposed to drugs that mimic or block natural hormones. Compared to these groups, our sample is more representative of the general population. In addition, since all of our children’s mothers had undergone amniocentesis, whatever may be unusual about that population will be shared by all the participants. It is unlikely that aT levels are different in mothers who undergo amniocentesis compared to those that do not, because within this group no relationship was seen between aT and maternal age, alpha-fetoprotein level, paternal age, or parental education level (Lutchmaya, 2000).

In conclusion, although we found little evidence that normal variation in prenatal testosterone levels is related to gender-typical play, we would be reluctant to dismiss prenatal hormone influences altogether, particularly considering the well-replicated findings in CAH. Instead, our study draws attention to the complexity of hormonal influences on behavior and highlights the need to consider dose and timing of exposure.

References

Abramovich, D. R., & Rowe, P. (1973). Foetal plasma testosterone levels at mid-pregnancy and at term: relationship to foetal sex. *Journal of Endocrinology, 56*, 621-622.

Abramovitch, D. R., Corter, C., Pepler, D. J., & Stanhope, L. (1986). Sibling and peer interaction: a final follow-up and comparison. *Child Development, 57*, 217-229.

Abramovitch, D. R., & Page, K. R. (1972). Pathways of water exchange in the fetoplacental unit at midpregnancy. *The Journal of Obstetrics and Gynaecology for the British Commonwealth, 79*, 1099.

Altman, D. G. (1991). *Practical Statistics for Medical Research*. London: Chapman and Hall.

Bates, J. E., & Bentler, P. M. (1973). Play activities of normal and effeminate boys. *Developmental Psychology, 9*, 20-27.

Beck-Peccoz, P., Padmanabhan, V., Baggiani, A. M., Cortelazzi, D., Buscaglia, M., Medri, G., et al. (1991). Maturation of hypothalamic-pituitary-gonadal function in normal human fetuses: circulating levels of gonadotropins, their common alpha-subunit and free testosterone, and discrepancy between immunological and biological activities of circulating follicle-stimulating hormone. *Journal of Clinical Endocrinology and Metabolism, 73*, 525-532.

Berenbaum, S. A. (2001). Cognitive Function in Congenital Adrenal Hyperplasia. *Endocrinology and Metabolism Clinics of North America, 30*, 173-191.

Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science, 3*, 203-206.

Berenbaum, S. A., Duck, S. C., & Bryk, K. (2000). Behavioral effects of prenatal *versus* postnatal androgen excess in children with 21-hydroxylase deficient congenital adrenal hyperplasia. *The Journal of Clinical Endocrinology and Metabolism, 85*, 727-733.

Berenbaum, S. A., & Snyder, E. (1995). Early hormonal influences on childhood sex-typed activity and playmate preferences: Implications for the development of sexual orientation. *Developmental Psychology, 31*, 31-42.

Birke, L. I., & Sadler, D. (1983). Progestin-induced changes in play behaviour of the prepubertal rat. *Physiology and Behavior, 30*, 341-347.

Brace, R. A., & Resnik, R. R. (1989). Dynamics and disorders of amniotic fluid. In R. K. Creasy & R. Resnik (Eds.), *Maternal-fetal medicine: Principles and practice*. Philadelphia: Harcourt Brace Jovanovich.

Carson, D. J., Okuno, A., Lee, P. A., Stetten, G., Didolkar, S. M., & Migeon, C. J. (1982). Amniotic fluid steroid levels. Fetuses with adrenal hyperplasia, 46, XXY fetuses, and normal fetuses. *American Journal of Disease in Childhood, 136*, 218-222.

Casto, J. M., Ward, O. B., & Bartke, A. (2003). Play, copulation, anatomy, and testosterone in gonadally intact male rats prenatally exposed to flutamide. *Physiology and Behavior, 79*, 633-641.

Chemes, H. E. (2001). Infancy is not a quiescent period of testicular development. *International Journal of Andrology, 24*, 2-7.

Clemens, L. G. (1974). Neurohormonal control of male sexual behavior. In W. Montagna & W. A. Sadler (Eds.), *Reproductive Behavior* (pp. 22-53). New York: Plenum Press.

Cole-Harding, S., Morstad, A. L., & Wilson, J. R. (1988). Spatial ability in members of opposite-sex twin pairs (abstract). *Behavior Genetics, 18*, 710.

Corder, J., & Stephan, C. (1984). Females' contribution of work and family roles: Adolescent aspiration. *Journal of Marriage and the Family, 56*, 391-402.

De Vries, G. J., & Simerly, R. B. (2002) Anatomy, development, and function of sexually dimorphic neural circuits in the mammalian brain. In D. W. Pfaff, A. P. Arnold, A. Etgen, S. Fahrbach & R. T. Rubin (Eds.), *Hormones, Brain and Behavior* (Vol IV, pp. 137-191). New York: Plenum Press.

DiPietro, J. A. (1981). Rough and tumble play: A function of gender. *Developmental Psychology, 17*, 50-58.

Dittman, R. W., Kappes, M. H., Kappes, M. E., Borger, D., Stegner, H., Willig, R. H., et al. (1990). Congenital adrenal hyperplasia I: Gender related behaviour and attitudes in female patients and sisters. *Psychoneuroendocrinology, 15*, 401-420.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist, 50*, 145-158.

Ehrhardt, A. A., Epstein, R., & Money, J. (1968). Fetal androgens and female gender identity in the early treated adrenogenital syndrome. *Johns Hopkins Medical Journal, 122*, 160-167.

Fels, E., & Bosch, L. R. (1971). Effect of prenatal administration of testosterone on ovarian function in rats. *American Journal of Obstetrics and Gynecology, 111*, 964-969.

Findlay, A. L. R. (1984). *Reproduction and the fetus*. Baltimore: University Park.

Finegan, J. A., Sitarenios, G., Bolan, P. L., & Sarabura, A. D. (1996). Children whose mothers had second trimester amniocentesis: follow up at school age. *British Journal of Obstetrics and Gynaecology, 103*, 214-218.

Finegan, J. K., Bartleman, B., & Wong, P. Y. (1989). A window for the study of prenatal sex hormone influences on postnatal development. *The Journal of Genetic Psychology, 150*, 101-112.

Forest, M. G., Betuel, H., Couillin, P., & Boue, A. (1981). Prenatal diagnosis of congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency by steroid analysis in the amniotic fluid of mid-pregnancy: comparison with HLA typing in 17 pregnancies at risk for CAH. *Prenatal Diagnosis, 1*, 197-207.

Forest, M. G., & Cathiard, A. M. (1975). Pattern of plasma testosterone and ^^{4}-androstenedione in normal newborns: Evidence for testicular activity at birth. *Journal of Clinical Endocrinology and Metabolism, 41*, 977-980.

Forest, M. G., Sizonenko, P. C., Cathiard, A. M., & Bertrand, J. (1974). Hypophyso-gonadal function in humans during the first year of life: I. Evidence for testicular activity in early infancy. *The Journal of Clinical Investigation, 53*, 819-828.

Goy, R. W. (1978). Development of play and mounting behaviour in female rhesus virilized prenatally with esters of testosterone and dihydrotestosterone. In D. J.

Chivers & J. Herbert (Eds.), *Recent Advances in Primatology. I. Behaviour* (pp. 449-462). London: Academic Press.

Goy, R. W. (1981). Differentiation of male social traits in female rhesus macaques by prenatal treatment with androgens: Variation in the type of androgen, duration, and timing of treatment. In M. J. Novy & J. A. Resko (Eds.), *Fetal Endocrinology* (pp. 319-339). New York: Academic Press.

Goy, R. W., Bercovitch, F. B., & McBrair, M. C. (1988). Behavioral masculinization is independent of genital masculinization in prenatally androgenized female rhesus macaques. *Hormones and Behavior, 22*, 552-571.

Greif, E. B. (1976). Sex role playing in pre-school children. In J. S. Bruner, A. Jolly & K. Sylva (Eds.), *Play-Its Role in Development and Evolution* (pp. 385-393). Harmondsworth: Penguin Books Ltd.

Grimshaw, G. M., Bryden, M. P., & Finegan, J. K. (1995). Relations between prenatal testosterone and cerebral lateralization in children. *Neuropsychology, 9*, 68-79.

Grimshaw, G. M., Sitarenios, G., & Finegan, J. (1995). Mental rotation at 7 years: relations with prenatal testosterone levels and spatial play experiences. *Brain and Cognition, 29*, 85-100.

Hany, J., Lilienthal, H., Sarasin, A., Roth-Harer, A., Fastabend, A., Dunemann, L., et al. (1999). Developmental exposure of rats to a reconstituted PCB mixture of Arclor 1254: Effects on organ weights, aromatase activity, sex hormone levels, and sweet preference behavior. *Toxicology and Applied Pharmacology, 158*, 231-243.

Henderson, B. A., & Berenbaum, S. A. (1997). Sex-typed play in opposite-sex twins. *Developmental Psychobiology, 31*, 115-123.

Hines, M. (2002). Sexual differentiation of human brain and behavior. In D. W. Pfaff, A. P. Arnold, A. Etgen, S. Fahrbach & R. T. Rubin (Eds.), *Hormones, Brain and Behavior* (Vol. IV, pp. 425-462). New York: Academic Press.

Hines, M., Ahmed, S. F., & Hughes, I. A. (2003). Psychological outcomes and gender-related development in complete androgen insensitivity syndrome. *Archives of Sexual Behavior, 32*, 93-101.

Hotchkiss, A. K., Ostby, J. S., Vandenbergh, J. G., & Gray, L. E. (2003). An environmental antiandrogen, vinclozolin, alters the organization of play behavior. *Physiology and Behavior, 79*, 151-156.

Hughes, F. P. (1991). Gender differences in play. In *Children, Play, & Development* (Second Edition ed., pp. 121-141). Boston: Allyn and Bacon.

Humphreys, A. P., & Smith, P. K. (1984). Rough-and-tumble play in preschool and playground. In P. K. Smith (Ed.), *Play in Animals and Humans* (pp. 241-270). Oxford: Blackwell.

Kaya, H., Hany, J., Fastabend, A., Roth-Harer, A., Winneke, G., & Lilienthal, H. (2002). Effects of maternal exposure to a reconstituted mixture of polychlorinated biphenyls on sex-dependent behaviors and steroid hormone concentrations in rats: Dose-response relationship. *Toxicology and Applied Pharmacology, 178*, 71-81.

Klopper, A. (1970). Steroids in amniotic fluid. *Annals of Clinical Research, 2*, 289-299.

Knickmeyer, R. C., Baron-Cohen, S., Raggatt, P., & Taylor, K. (in press). Foetal testosterone, social relationships, and restricted interests in children. *Journal of Child Psychology, Psychiatry, and Allied Disciplines*.

Kulik, L. (2002). The impact of social background on gender-role ideology: Parents' versus children's attitudes. *Journal of Family Issues, 23*, 53-73.

Lackey, P. N. (1989). Adult attitudes about assignments of household chores to male and female children. *Sex Roles, 20*, 271-282.

Leveroni, C. L., & Berenbaum, S. A. (1998). Early androgen effects on interest in infants: evidence from children with congenital adrenal hyperplasia. *Development Neuropsychology, 14*, 321-340.

Lippa, R., & Hershberger, S. (1999). Genetic and environmental influences on individual differences in masculinity, femininity, and gender diagnosticity: Analyzing data from a classic twin study. *Journal of Personality, 67*, 127-155.

Lutchmaya, S. (2000). *Foetal testosterone and social-communicative development.* Unpublished Ph.D. Dissertation, University of Cambridge, Cambridge, UK.

Lutchmaya, S., Baron-Cohen, S., & Raggatt, P. (2002a). Foetal testosterone and vocabulary size in 18- and 24-month-old infants. *Infant Behavior and Development, 24*, 418-424.

Lutchmaya, S., Baron-Cohen, S., & Raggett, P. (2002b). Foetal testosterone and eye contact in 12 month old infants. *Infant Behavior and Development, 25*, 327-335.

Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: a meta-analysis. *Psychological Bulletin, 109*, 267-296.

MacLusky, N., & Naftolin, F. (1981). Sexual differentiation of the central nervous system. *Science, 211*, 1294-1303.

Meaney, M. J., & McEwen, B. S. (1986). Testosterone implants into the amygdala during the neonatal period masculinise the social play of juvenile female rats. *Brain Research, 398*, 324-328.

Meisel, R. L., & Ward, I. L. (1981). Fetal female rats are masculinized by male littermates located caudally in the uterus. *Science, 213*, 239-242.

Meyer-Bahlburg, H. F. L., Erhardt, A. A., & Feldman, J. F. (1985). Questionnaires for the assessment of atypical gender role behavior: A methodological study. *Journal of the American Academy of Child Psychiatry, 24*, 695-701.

Meyer-Bahlburg, H. F. L., Feldman, J. F., Cohen, P., & Ehrhardt, A. A. (1988). Perinatal factors in the development of gender-related play behavior: Sex hormones versus pregnancy complications. *Psychiatry, 51*, 260-271.

Meyer-Bahlburg, H. F. L., Sandberg, D. E., Dolezal, C. L., & Yager, T. J. (1994). Gender-related assessment of childhood play. *Journal of Abnormal Child Psychology, 22*, 643-660.

Nagamani, M., McDonough, P. G., Ellegood, J. O., & Mahesh, V. B. (1979). Maternal and amniotic fluid steroids throughout human pregnancy. *American Journal of Obstetrics and Gynecology, 134*, 674-680.

New, M. I. (2003). Inborn errors of adrenal steroidogenesis. *Molecular and Cellular Endocrinology, 211*, 75-83.

Nordenstrom, A., Servin, A., Bohlin, G., Larsson, A., & Wedell, A. (2002). Sex-typed toy play behavior correlates with the degree of prenatal androgen exposure assessed by the *CYP21* genotype in girls with congenital adrenal hyperplasia. *The Journal of Clinical Endocrinology and Metabolism, 87*, 5119-5124.

Pang, S., Levine, L. S., Cedreqvist, L. L., Fuentes, M., Riccardi, V. M., Holcombe, J. H., et al. (1980). Amniotic fluid concentrations of delta 5 and delta 4 steroids in fetuses with congenital adrenal hyperplasia due to 21-hydroxylase deficiency and in anencephalic fetuses. *Journal of Clinical Endocrinology and Metabolism, 51*, 223-229.

Pang, S., Levine, L. S., Chow, D. M., Faiman, C., & New, M. I. (1979). Serum androgen concentrations in neonates and young infants with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Clinical Endocrinology, 11*, 575-584.

Pellegrini, A. D., & Smith, P. K. (1998). Physical activity play: the nature and function of a neglected aspect of play. *Child Development, 69*, 577-598.

Pellis, S. M. (2002). Sex differences in play fighting revisited: Traditional and nontraditional mechanisms of sexual differentiation in rats. *Archives of Sexual Behavior, 31*, 17-26.

Phoenix, C. H., Goy, R. W., Gerall, A. A., & Young, W. C. (1959). Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology, 65*, 369-382.

Powlishta, K. K., Sen, M. G., Serbin, L. A., Poulin-Dubois, D., & Eichstedt, J. A. (2001). From infancy through middle childhood: The role of cognitive and social factors in becoming gendered. In R. K. Unger (Ed.), *Handbook of the Psychology of Women and Gender* (pp. 116-132). New York: John Wiley & Sons.

Quadagno, D. M., Briscoe, R., & Quadagno, J. S. (1977). Effects of perinatal gonadal hormones on selected nonsexual behavior patterns: A critical assessment of the nonhuman and human literature. *Psychological Bulletin, 84*, 62-80.

Quarm, D. (1983). The effect of gender on sex-role attitudes. *Sociological Focus, 16*, 285-303.

Resnick, S. M., Gottesman, I. I., & McGue, M. (1993). Sensation seeking in opposite-sex twins: an effect of prenatal hormones? *Behavior Genetics, 23*, 323-329.

Reyes, F. I., Boroditsky, R. S., Winter, J. S. D., & Faiman, C. (1974). Studies on human sexual development. II. Fetal and maternal serum gonadotropin and sex steroid concentrations. *Journal of Clinical Endocrinology and Metabolism, 38*, 612-617.

Robinson, J. D., Judd, H. L., Young, P. E., Jones, D. W., & Yen, S. S. C. (1977). Amniotic fluid androgens and estrogens in midgestation. *Journal of Clinical Endocrinology, 45*, 755-761.

Rodgers, C. S., Fagot, B. I., & Winebarger, A. (1998). Gender-typed toy play in dizygotic twin pairs: a test of hormone transfer theory. *Sex Roles, 39*, 173-184.

Servin, A., Nordenstrom, A., Larsson, A., & Bohlin, G. (2003). Prenatal androgens and gender-typed behavior: A study of girls with mild and sever forms of congenital adrenal hyperplasia. *Developmental Psychology, 39*, 440-450.

Smith, L. L., & Hines, M. (2000). Language lateralization and handedness in women prenatally exposed to diethylstilbestrol (DES). *Psychoneuroendocrinology, 25*, 497-512.

Stoneman, Z., Brody, G. H., & MacKinnon, C. E. (1986). Same-sex and cross-sex siblings: Activity choices, roles, behavior, and gender stereotypes. *Sex Roles, 15*, 495-511.

Togeby, L. (1995). Feminist attitudes: Social interests or political ideology. *Women and Politics, 15*, 424-432.

Udry, J. R., Morris, N. M., & Kovenock, J. (1995). Androgen effects on women's gendered behaviour. *Journal of Biosocial Science, 27*, 359-368.

Vreugdenhil, H. J. I., Slijper, F. M. E., Mulder, P. G. H., & Weisglas-Kuperus, N. (2002). Effects of perinatal exposure to PCBs and dioxins on play behavior in Dutch children at school age. *Environmental Health Perspectives, 110,* A593-A598.

Wallen, K., & Baum, M. J. (2002). Masculinization and defeminization in altricial and precocial mammals: Comparative aspects of steroid hormone action. In D. W. Pfaff, A. P. Arnold, A. Etgen, S. Fahrbach & R. T. Rubin (Eds.), *Hormones, Brain and Behavior* (Vol. IV, pp. 385-423). New York: Academic Press.

Warne, G. L., Faiman, C., Reyes, F. I., & Winter, J. S. (1977). Studies on human sexual development. V. Concentrations of testosterone, 17-hydroxyprogesterone and progesterone in human amniotic fluid throughout gestation. *Journal of Clinical Endocrinology and Metabolism, 44*(5), 934-938.

Wathen, N. C., Campbell, D. J., Kitau, M. J., & Chard, T. (1993). Alphafetoprotein levels in amniotic fluid from 8 to 18 weeks of pregnancy. *British Journal of Obstetrics and Gynaecology, 100*, 380-382.

White, P. C., & Speiser, P. W. (2002). Long-term consequences of childhood-onset congenital adrenal hyperplasia. *Best Practice & Research Clinical Endocrinology and Metabolism, 16*, 273-288.

Young, W. C., Goy, R. W., & Phoenix, C. H. (1964). Hormones and sexual behaviour. *Science, 143*, 212-218.

Zucker, K. J., Bradley, S. J., Oliver, G., Blake, J., Fleming, S., & Hood, J. (1996). Psychosexual development of women with congenital adrenal hyperplasia. *Hormones and Behavior, 30*, 300-318.

Appendix A

Pilot Study of the Modified Child Game Participation Questionnaire

Participants

Participants were 28 girls and 35 boys from local schools in Cambridgeshire. All children were 4 to 5 years old and attended reception/nursery classes. This age was chosen to match that of the children in our group's longitudinal study of prenatal testosterone and child development. Children were given the questionnaire and a letter explaining the study in sealed envelopes at school and told to take them home to their parents. If parents agreed to participate they completed the questionnaire and returned it by freepost envelope. The 63 participating children represent those whose parents responded from a larger sample of 80.

The Children's Play Questionnaire

The questionnaire was adapted from the Child Game Participation Questionnaire (CGPQ) (Bates & Bentler, 1973). This instrument was originally developed to discriminate between boys with gender identity disorder and gender-typical boys. It also shows highly significant gender differences (Meyer-Bahlburg et al., 1985). The original item pool included 120 children's games. In order to make the questionnaire more manageable and increase the response rate, it was decided to shorten the questionnaire. In addition, given the young age of the sample, some items, such as softball, were not appropriate. The final questionnaire included 10 items that were expected to be preferred by boys and 10 that were expected to be preferred by girls. 9 of the male items appeared in the bipolar Gender scale as masculine items in Meyer-Bahlburg, Sandberg, Dolezal, and Yager's (1994) factor-analytic study of a modified CGPQ. The difference between boys and girls on this factor showed a large effect size, $d = 3.90$. We added one item ("playing with blocks or lego/duplo") that seemed gender dimorphic in our region. 9 of the female items appeared in the bipolar Gender scale as feminine items in Meyer-Bahlburg et al.'s (1994) study. We added one item ("playing with hair") that seemed gender dimorphic in our region. We split the item "plays with dolls" into 2 items: "playing with Barbie-type dolls" and "playing with baby dolls" in order to have equal numbers of male and female items. The questionnaire mailed to parents also included 10 items thought to be gender neutral. This was done to prevent biased answering, which might have occurred if parents realized the test focused specifically on gender-typical play.

For each game, parents indicated their child's interest on a Likert scale where 1 was not at all interested and 5 was very interested.

Results

Two scores were calculated for each individual. A total femininity score was calculated by adding together the score on each female item (1=0, 2=1, 3=2, 4=3, 5=4). A total masculinity score was calculated by adding together the score on each male item in the same way. The femininity and masculinity scores had a possible range of 0 to 40. Table A1 shows descriptive data for each scale by gender. Both scales showed significant differences and large effect sizes. Effect sizes are similar to that reported for the composite scale on the modified CGPQ used by Meyer-Bahlburg et al. (1994) with 6-10 year old children.

T-tests were run on every item in order to determine whether they showed the expected sex-differences (t-tests were also run on neutral items to ensure that they did not show a sex-difference). Equal variances were not assumed. The probability of a type I error was maintained at 0.05 for all analyses. Because this involved 30 comparisons, it was necessary to use a Bonferroni correction when evaluating the results. Sex-differences were considered significant if they had a t value greater than 3.32. Table A2 shows an item analysis for the pilot questionnaire. 3 of the male items did not show significant sex-differences although they showed trends in the expected direction (Building play houses, forts, huts or dens; Playing with blocks or Lego/Duplo; Climbing trees/rope ladders). All of the female items showed significant sex-differences in the correct direction. We decided to eliminate the three male items from the next version of the test. In order to keep the number of male and female equivalent we split the items “pretending to be a soldier or a super-hero,” “Playing Cowboys and Indians or similar/play-fighting” and “Playing with toy vehicles (eg cars, trucks, planes, trains)” into 2 separate items each. 2 of the neutral items (“using coloring books” and “doing arts and crafts/painting” showed significant sex-differences in a female direction. These were eliminated from the next version of the questionnaire.

References

Bates, J. E., & Bentler, P. M. (1973). Play activities of normal and effeminate boys. *Developmental Psychology, 9*, 20-27.

Meyer-Bahlburg, H. F. L., Erhardt, A. A., & Feldman, J. F. (1985). Questionnaires for the assessment of atypical gender role behavior: A methodological study. *Journal of the American Academy of Child Psychiatry, 24,* 695-701.

Meyer-Bahlburg, H. F. L., Sandberg, D. E., Dolezal, C. L., & Yager, T. J. (1994). Gender-related assessment of childhood play. *Journal of Abnormal Child Psychology, 22*, 643-660.

Appendix B

Hormone Assays

Testosterone

Amniotic fluid was extracted with diethyl ether. Recovery experiments have demonstrated 95 percent recovery of testosterone using this method. The ether was evaporated to dryness at room temperature and the extracted material redissolved in assay buffer. The testosterone was assayed by the DPC ‘Count-a-Coat’ method (Diagnostic Products Corp, Los Angeles, CA 90045-5597), which uses an antibody to testosterone coated onto propylene tubes and a 125-I labeled testosterone analogue. The detection limit of the assay is approximately 0.1 nmol/L. Intra-assay coefficients of variation (i.e. 1 standard deviation expressed as a percentage of the mean value) were between 10 and 15%. This method measures total extractable testosterone.

Estrogen

Amniotic fluid was extracted with diethyl ether. Recovery experiments have demonstrated 95 percent recovery of estradiol using this method. The estradiol was measured by fluorescence-labeled immunoassay. The Wallac-Delfia method was used (Wallac OY, Turku, Finland). This assay uses a polyclonal rabbit antibody to estradiol in a competitive format in which sample estradiol competes with europium-labeled estradiol analogue for the antibody binding sites. A second antibody directed against rabbit IgG is coated to the microtitre plate and is used to capture the first antibody and its bound estradiol analogue. After washing, the europium is measured by time-resolved fluorescence. Calibration is with pure 17beta-estradiol. The detection limit is 25 pmol/L. The cross reactivity with steroids other than 17beta estradiol is very low. It should be noted that 16 hydroxy and 16 oxo-steroids, steroids that are formed in the feto-placental unit, cross react to less than 0.9% by weight. Intra-assay coefficients of variation were 5.2% at 180 pmol/L and 3.9% at 875 pmol/L.

Alpha-fetoprotein

AFP was measured by fluorescence-labeled immunoassay. The Wallac-Delfia method was used (Wallac OY, Turku, Finland). This assay is based on the direct sandwich technique in which two monoclonal antibodies (derived from mice) are directed against two separate antigenic determinants on the AFP molecule. The analytical sensitivity of the assay is typically better than 0.1 U/ml. Recovery experiments have demonstrated 101 percent recovery of AFP using this method. Serum albumin concentrations in the normal physiological range do not interfere with AFP determination. Intra-assay coefficients of variation were 1.0% at 10199 U/mL and 1.1 at 12438 U/mL.

Author Note

The reported study was approved by the Eastern Multi-regional Ethics Committee and all relevant local ethics committees. Informed consent was obtained in all cases.

We are grateful to the Nancy Lurie Marks Family Foundation for supporting this work. RK was also supported by a British Government Overseas Research Studentship (ORS) and the Cambridge Overseas Trust. This work was submitted in part fulfillment of the degree of PhD, University of Cambridge by RK. SW and SB were also supported by the Medical Research Council (MRC) UK. We are grateful to Ian Goodyer, Joe Herbert, Svetlana Lutchmaya, and Rosa Hoekstra for their support and discussion. Finally we are indebted to the families who participated in this study.

Reprint enquires should be addressed to email address.

Table 1

Mean, Standard Deviation and Range for Outcome and Predictor Variables by Sex

	Girls n = 22			Boys n = 31			
Variable	Mean	SD	Range	Mean	SD	Range	*d*
aT(nmol/L)**	0.39	0.18	0.17-0.80	1.02	0.44	0.13-1.80	1.87
Estrogen (pmol/L)	938	373	496-1950	972	420	440-2630	0.08
AFP (MU/L)	11.1	3.15	6.50-19.7	11.2	4.12	3.10-22.0	0.02
Gestational Age at Amnio	16.8	0.97	14-18	16.8	1.79	14-20	0.20
Maternal Age	34.2	4.63	23-40	35.6	5.01	25-42	0.28
Maternal Education	3.21	0.89	2-4	3.27	0.87	2-5	0.06
Male Items**	7.61	6.04	0-20	28.45	5.41	18-40	3.63
Female Items**	30.7	7.14	16-40	10.7	4.89	1-19	3.27
Neutral Items*	32.7	3.96	22.5-40	29.4	4.50	20-37.5	0.77

$^{*}p < 0.05$, $^{**}p < 0.01$

Table 2

Summary of Hierarchical Regression Analyses Testing the Contribution of Sex and aT to Scores on Male and Female Items

Variable	*B*	*SE B*	β
Male Items			
Step 1			
Sex	-20.8	1.63	-0.87**
Step 2			
Sex	-21.2	2.23	-0.89**
aT	-0.58	2.30	-0.02
Step 3			
Sex	-17.9	4.22	-0.75**
aT	3.70	4.17	-0.15
Interaction of Sex and aT	3.75	4.17	-0.24
Female Items			
Step 1			
Sex	19.6	1.70	0.85**
Step 2			
Sex	20.4	2.30	0.89**
aT	1.22	2.38	0.05
Step 3			
Sex	13.3	4.23	0.58**

aT	8.12	4.19	0.34
Interaction of Sex and aT	-8.28	4.19	-0.57

Note. For Male Items: $R^2 = 0.76$ for Step 1 ($p < 0.01$); $\Delta R^2 = 0.00$ for Step 2 (*ns*); $\Delta R^2 = 0.00$ for Step 3 (*ns*). For Female Items: $R^2 = 0.72$ for Step 1 ($p < 0.01$); $\Delta R^2 = 0.00$ for Step 2 (*ns*); $\Delta R^2 = 0.02$ for Step 3 (*ns*).

$^{*}p < 0.05$, $^{**}p < 0.01$

Table 3

Correlation matrix showing relationships between the independent and dependent variables for all participants of both sexes (n = 40-53)

Variable	Male Items	Female Items	Sex	aT	Estrogen[a]	AFP	Gest. Age	Mat Age	Mat Edu	Older Bro	Older Sis
Male Items											
Female Items	-0.75**										
Sex	-0.87**	0.85**									
aT	0.57**	-0.54**	-0.67**								
Estrogen[a]	-0.01	0.05	-0.05	0.29							
AFP	0.10	-0.04	-0.01	0.11	0.35*						
Gest. Age	0.07	-0.13	-0.10	0.16	-0.10	-0.57**					
Mat. Age	0.27*	-0.20	-0.16	0.15	-0.14	0.18	-0.50**				
Mat. Edu	-0.04	-0.19	-0.02	-0.12	0.03	0.06	-0.33	0.34*			
Older Bro	0.19	-0.16	-0.11	0.13	-0.16	0.24	-0.05	0.22	0.01		
Older Sis	0.07	-0.12	-0.14	-0.04	-0.24	0.13	-0.30	0.33	0.29	0.13	

Note. n varies from 40-53 due to missing data from some participants

[a] Represents log(estrogen)

[*] $p < 0.05$, [**] $p < 0.01$

Table 4

Final Model: Hierarchical Regression Incorporating Background Variables: Maternal Age and Gestational Age

Variable	*B*	*SE B*	β
Male Items			
Gestational Age	0.38	0.75	0.05
Maternal Age	0.33	0.22	0.14
Sex	-20.4	1.84	-0.85**
Female Items			
Gestational Age	-0.78	0.76	-0.10
Maternal Age	-0.25	0.22	-0.11
Sex	18.3	1.87	0.83**

$^{*}p < 0.05$, $^{**}p < 0.01$

Table A1

Scores of 4-5 Year Old Children on Female and Male Items on the Pilot Version of the Children's Play Questionnaire

	Girls n = 28			Boys n = 35			
Scale	Mean	SD	Range	Mean	SD	Range	*d*
Female Items**	31.37	5.69	15-39	11.12	5.68	0-21	3.57
Male Items**	11.89	4.85	4-23	28.22	6.91	7-39	2.74

** $p < 0.01$

Table A2

Mean Scores of 4-5 Year Old Children by Sex for All Items on the Pilot Version of the Children's Play Questionnaire

Item	Sex	Mean	*t*-value
Playing with Barbie-type dolls*	Female Male	4.00 1.50	8.56
Role-playing domestic activities (e.g. cooking, cleaning, bathing)*	Female Male	4.50 3.00	6.49
Playing dress up (fashion/jewelry)*	Female Male	4.71 2.40	8.99
Role-playing family relationships (e.g. parenting/marriage)*	Female Male	4.30 2.46	6.49
Skipping rope or skipping*	Female Male	3.25 1.69	5.81
Playing school (pretending to be a teacher)*	Female Male	4.25 2.34	5.90
Dancing*	Female Male	4.64 2.77	6.56
Playing with hair (e.g. brushing someone else's hair)*	Female Male	3.54 1.51	7.61
Playing tea-parties*	Female Male	3.89 2.20	6.31
Playing with baby dolls*	Female Male	4.14 1.49	12.29
Pretending to be a soldier or super-hero*	Female Male	1.29 3.91	-10.58

Playing with toy guns or other weapons*	Female Male	1.46 3.88	-8.28
Playing with toy vehicles (e.g. cars, trucks, planes, trains)*	Female Male	2.54 4.42	-7.47
Pretending to be an astronaut (space-man) or explorer*	Female Male	1.36 3.46	-8.41
Playing with toy tools*	Female Male	2.18 3.66	-4.87
Playing with electric trains*	Female Male	2.11 3.71	-4.84
Playing Cowboys and Indian or similar/play-fighting*	Female Male	1.29 3.69	-8.59
Building play houses, forts, huts, or dens	Female Male	3.46 3.89	-1.51
Playing with blocks or Lego/Duplo	Female Male	3.50 4.03	-2.27
Climbing trees/rope ladders	Female Male	3.04 4.00	-3.21
Looking at picture books	Female Male	4.61 4.41	.975
Using colouring books*	Female Male	4.59 3.69	3.74
Playing with stuffed animals	Female Male	3.82 2.91	2.94
Riding on tricycles/bicycles	Female Male	4.57 4.77	-1.29
Swimming	Female Male	4.46 4.06	1.66

Playing on swings	Female Male	4.54 4.23	1.42
Playing on see-saws	Female Male	3.43 3.71	-1.04
Doing arts and crafts/painting*	Female Male	4.75 4.09	3.49
Watching cartoons	Female Male	3.79 4.26	-1.73
Playing board-games (e.g. Ludo, Snakes and Ladders)	Female Male	3.61 3.49	.443

Note. Raw scores were used for each item (lowest score =1, highest=5)

Note. First 10 items were expected to show a female preference, the next 10 items were expected to show a male preference, the last 10 items were expected to show no preference

* $p < 0.05$, using a Bonferroni correction