Journal of Child Psychology and Psychiatry 46:2 (2005), pp 198–210 doi: 10.1111/j.1469-7610.2004.00349.x

Foetal testosterone, social relationships, and restricted interests in children

Rebecca Knickmeyer,[1] Simon Baron-Cohen,[1] Peter Raggatt,[2] and Kevin Taylor[2]
[1]Autism Research Centre, Departments of Experimental Psychology and Psychiatry, University of Cambridge, UK;
[2]Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge, UK

Background: Sex-differences exist in some areas of human social behaviour. In animals, foetal testosterone (fT) plays a central role in organising the brain and in later social behaviour. fT has also been implicated in language development, eye-contact, and spatial ability in humans. **Methods:** Fifty-eight children (35 male and 23 female), whose fT was analysed in amniotic fluid, were followed up at age 4. Their mothers completed the Children's Communication Checklist, a questionnaire assessing language, quality of social relationships and restricted interests. **Results:** fT was negatively correlated to quality of social relationships, taking sex-differences into account. fT was also positively correlated with restricted interests in boys. **Conclusions:** These findings implicate fT in both social development and attentional focus. They may also have implications for understanding the sex ratio in autism. **Keywords:** Testosterone, social relationships, restricted interests, sex-differences, autism.
Abbreviations: fT: foetal testosterone.

Experiments in non-human mammals show that foetal testosterone (fT) plays a critical role in a wide range of sex-differences, although the contribution of oestrogen of ovarian origin (Fitch & Bimonte, 2002; Fitch & Denenberg, 1998) and direct genetic influences on sexual differentiation (Arnold, 1996; De Vries et al., 2002) are increasingly recognised. Foetal testosterone (fT) has been shown to affect the anatomy of the brain, including the hypothalamus, limbic system, and neocortex (Arnold & Gorski, 1984; Breedlove, 1994; MacLusky & Naftolin, 1981), sexually dimorphic behaviours such as aggression and activity level (Goy & McEwen, 1980) and sexually dimorphic cognitive abilities such as spatial navigation (Williams & Meck, 1991). In general, these animal experiments have compared castrated males, normal males, normal females, and females treated with androgen. Castrated males usually show feminised neural development, cognition, and behaviour, while females treated with androgen show masculinised neural development, cognition, and behaviour.

In human beings, sex-differences are apparent both in brain structures and in cognitive skills (Breedlove, 1994; Collaer & Hines, 1995; Halpern, 1992; Kimura, 1999; MacLusky & Naftolin, 1981; Baron-Cohen, 2003). The psychological study of sex-differences has traditionally focused on spatial, mathematical, and verbal ability (Kimura, 1999). However, there is increasing interest in potential sex-differences in social relationships. The quality of social relationships depends on a variety of abilities. Based on evolutionary arguments, Geary (1998, 2002) suggests three socio-cognitive abilities that should show a female superiority: the ability to read nonverbal communication signals (i.e., body posture and facial expressions), language, and theory of mind. Baron-Cohen (2002, 2003) proposes that females, on average, are better at 'empathising'. He defines this as the drive to identify another's mental states and respond to these with an appropriate emotion. This encompasses what is referred to as using a theory of mind but includes an affective reaction as well.

Several studies have shown a female advantage in reading nonverbal signals. A meta-analytic study by Hall (1978, 1984) showed that females were on average better than males at interpreting body language, vocal tone, and facial expression. In a more recent study by Baron-Cohen, Jolliffe, Mortimore, and Robertson (1997), women were better at attributing subtle mental states to a person, when interpreting the eye region of the face. However, not all studies show this effect (Gitter, Black, & Mostofsky, 1972). Part of this variation may depend on the specific emotions being examined. For example, a study by Rotter and Rotter (1988) showed that while females were better at identifying emotions overall, males were superior to females at recognising male anger.

Although a female superiority for language-related skills is commonly accepted, actual results vary considerably across studies. This is not surprising given that language consists of a number of subsystems, including phonology, morphology, the lexicon, semantics, syntax, pragmatics, and discourse (Gleason & Ely, 2002). There are well-replicated female advantages for verbal memory, spelling ability and verbal fluency in adulthood, although females do not have a larger vocabulary than males (Kimura, 1999). Developmentally, a number of studies have reported greater vocabularies and faster rates of language acquisition in girls (Fenson et al., 1994; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Hyde & Linn, 1988; Reynell & Huntley, 1985).

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

Theory of mind is the ability to make inferences about the intentions, beliefs, and emotions of other people in order to predict and explain their behaviour. Research into sex-differences in theory of mind has been limited because many of the associated tests are not sensitive enough to detect subtle individual differences, including sex-differences (Baron-Cohen et al., 1997). There are several studies, though, that suggest that theory of mind may develop earlier in females and that girls and women are, on average, better at making inferences about people's mental states and adjusting their behaviour accordingly (Banerjee, 1997; Baron-Cohen et al., 1997; Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999; Happé, 1995). These differences may arise, in part, from differences in social interest. Young girls show a preference for dyadic interactions (Benenson, 1993) and are more interested in facial than spatial/mechanical stimuli even at birth (Connellan, Baron-Cohen, Wheelwright, Ba'tki, & Ahluwalia, 2001).

Sex-differences in social development can also be examined by looking at sex-biases in developmental conditions. Specific language delay, semantic-pragmatic disorder, and autism spectrum conditions are all more common in males (Bishop, 1990; Rutter, 1978; Wing, 1981). Autism in particular has been described as an extreme manifestation of some sexually dimorphic traits or an 'extreme male brain' (see Baron-Cohen, 2002). Individuals with autism perform poorly on tests where females are usually superior to males, such as the 'Reading the Mind in the Eyes' test (Baron-Cohen, Wheelwright, & Hill, 2001a), but perform better than people without autism on tests where males usually outperform females, such as the 'Embedded Figures Task' (Jolliffe & Baron-Cohen, 1997; Baron-Cohen & Hammer, 1997). Autism spectrum conditions are characterised by difficulties in social relationships, but also by narrow, restricted interests. If autism actually is an exaggeration of typical sex-differences, then normally developing males should show more restricted interests, compared to females. Is this the case?

Intriguing evidence comes from a study of the Autism-Spectrum Quotient (AQ). This self-administered instrument measures autistic traits, and consists of 5 subscales (communication, social, imagination, local details, and attention switching). Males score slightly but significantly higher on the scale as a whole and on all subscales except local details. Higher male scores on the imagination and attention switching scales suggest that males are more likely to have restricted interests than females (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001b). Another source of evidence comes from tests of the drive to understand systems ('systemising' see Baron-Cohen 2002, 2003), which, by definition, involves a very narrow focus of attention on each variable in the system. Males score higher on both the 'Systemising Quotient' (SQ) (Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003) and on an experimental test of systemising (Lawson, Baron-Cohen, & Wheelwright, 2004).

Social and environmental factors certainly play a significant role in the development of sex-differences in social development, but biological contributions are also suggested by genetic and neurobiological studies. Social skills, social reciprocity, and empathy have a heritable component (Constantino & Todd, 2000; Scourfield, Martin, Lewis, & McGuffin, 1999; Zahn-Waxler, Robinson, & Emde, 1992). Several brain regions appear to subserve social cognition, and theory of mind in particular. This neural network comprises the amygdala, superior temporal sulcus (STS), and the orbital and medial prefrontal cortex (Adolphs, 2001; Baron-Cohen et al., 2000). In addition, both specific language delay and autism are considered to have a neurobiological basis and a strong genetic component (Bishop, 2001; Stodgell, Ingram, & Hyman, 2001).

Given the important role of fT in sexually dimorphic behaviour in animals, could fT play a role in human sex-differences in social development? The effects of prenatal hormones on human sex-differences are difficult to study because it is patently unethical to manipulate hormone levels in human foetuses. However, several other research strategies are available. These include studies of individuals with genetic conditions that alter prenatal hormone levels such as congenital adrenal hyperplasia (CAH) (Berenbaum & Hines, 1992; Berenbaum & Snyder, 1995; Hampson, Rovet, & Altmann, 1998; Nass et al., 1987; Plante, Boliek, Binkiewicz, & Erly, 1996), individuals exposed to synthetic sex hormones in utero such as diethylstilbestrol (Hines & Sandberg, 1996; Hines & Shipley, 1984; Meyer-Bahlburg et al., 1995; Reinisch & Sanders, 1992; Schacter, 1994; Smith & Hines, 2000) and females who shared the uterus with an opposite-sex, dizygotic twin (Cole-Harding, Morstad, & Wilson, 1988; Henderson & Berenbaum, 1997; Miller, 1994; Resnick, Gottesman, & McGue, 1993; Rodgers, Fagot, & Winebarger, 1998). Testosterone has also been measured in umbilical cord blood (Jacklin, Maccoby, & Doering, 1983; Jacklin, Wilcox, & Maccoby, 1988), maternal blood during pregnancy (Hines et al., 2002), and in amniotic fluid (Finegan, Niccols, & Sitarenios, 1992; Grimshaw, Sitarenios, & Finegan, 1995a; Grimshaw, Bryden, & Finegan, 1995b; Lutchmaya, Baron-Cohen, & Raggatt, 2002a, 2002b). Each method makes certain assumptions and has particular limitations (Baron-Cohen, Knickmeyer, & Lutchmaya, 2004). In this study we used fT measured in amniotic fluid to study the relationships between prenatal hormones and postnatal behaviour in 4-year-old children.

Measuring fT via amniocentesis

Both male and female foetuses produce some testosterone. In males the main source is the testes. Females are exposed to small amounts of fT from the foetal adrenal glands and from the maternal adrenals, ovaries and fat (Geschwind & Galaburda, 1985; Martin, 1985). Testosterone can be measured in amniotic fluid collected during midtrimester amniocentesis (Finegan, Bartleman, & Wong, 1989). Testosterone is thought to enter the amniotic fluid via diffusion through the foetus' skin in early pregnancy, and later from foetal urination (Klopper, 1970; Nagami, McDonough, Ellegood, & Mahesh, 1979; Robinson, Judd, Young, Jones, & Yen, 1977). Although the exact correlation between testosterone levels in the foetal serum and the amniotic fluid is unknown, the maximal sex-difference in amniotic testosterone between males and females occurs between weeks 12 and 18, closely paralleling peak serum levels (Finegan et al., 1989). In animal models, the general critical period for sexual differentiation of the brain usually occurs when sex-differences in serum testosterone are highest (Smith & Hines, 2000). Therefore it is likely that this is an important period for sexual differentiation of the human brain as well. This does not mean it is the only period for differentiation. There is another testosterone surge in early neonatal life and again at puberty (MacLusky & Naftolin, 1981). Sexual differentiation under direct genetic control could also occur when no sex-differences in testosterone are apparent. In addition, within the period of fT-related sexual differentiation, different characteristics are influenced at somewhat different times (Smith & Hines, 2000).

The first study to use this methodology was carried out by Finegan et al. (1992). They reported relationships with language comprehension, classification abilities, counting, number facts, and block building, but the results were not consistent with the predictions of androgen theory. This may be because the abilities studied did not show a sex-difference in their, or other, samples (Tierney, Smith, Axworthy, & Ratcliffe, 1984). Later studies by the same group have produced results more consistent with predictions. At age eight, girls with higher levels of fT performed a mental rotation task faster than girls with lower levels (Grimshaw et al., 1995a). At age 10, girls with higher levels of fT showed a more masculine pattern of cerebral lateralisation (Grimshaw et al., 1995b). In the children who took part in the study reported in the present paper, fT was negatively related to vocabulary size (Lutchmaya et al., 2002a) and amount of eye contact (Lutchmaya et al., 2002b) at 12 months of age.

One of the strengths of the amniocentesis design is that it measures testosterone produced by the foetus during a period in which masculinisation of the brain is thought to occur. A limitation of research using this method is that a truly random sample cannot be collected, since one can only include in a study those individuals who have decided/been advised to have an amniocentesis due to late maternal age or other factors that increase the risk of foetal abnormality.

In the present study we examined the relationship between fT measured in fluid obtained at amniocentesis and scores on the children's communication checklist (CCC) (Bishop, 1998) in four-year-old children. This instrument measures speech, syntax, pragmatic language skills, quality of social relationships, and restricted interests. The speech scale measures the intelligibility of speech and could be considered a measure of articulation. The syntax scale assesses basic grammar. Hyde and Linn (1988) argue from the results of a meta-analytic study that the effect sizes for sex-differences in such basic language skills are so slight as to be nonexistent. Therefore we predicted no relation between these scales and fT. However, it should be noted that some authors suggest that in very young children, girls are superior in both areas (Kimura, 1999). There are no reported sex-differences on these scales of the CCC.

Pragmatic language skills have been relatively little studied in this age group. Pragmatic skills are intrinsically social. The ability to adapt one's speech to different listeners would be facilitated by using a theory of mind. In a sample of 6-year-old twins, girls scored higher on the pragmatic composite of the CCC when completed by teachers. When completed by parents, girls scored higher on one of the subscales of the pragmatic composite (rapport), but not on the entire composite (Bishop and Laws, personal communication). Therefore we predicted that higher fT levels would be associated with poorer scores on the pragmatic language scale, although we were aware that the pragmatic scale included multiple components and this might reduce associations with fT.

The CCC includes the quality of social relationships and restricted interest scales so as to determine whether a diagnosis of autism should be considered. Given that autism is characterised by social deficits coupled with restricted interests and occurs far more often in males, we predicted that testosterone would affect these scales as well. In the sample of 6-year-old twins, there was a significant sex-difference in restricted interests when parents completed the scale, but the scale was unreliable. No sex-differences were seen on either scale when teachers completed the questionnaire (Bishop and Laws, personal communication). However, the quality of social relationships scale is comparable to instruments such as the 'Social Cognitive Skills Questionnaire' (Scourfield et al., 1999; Skuse et al., 1997), which shows a small but consistent sex-difference. Sex-differences in restricted interests are suggested by studies of the AQ and SQ, as described earlier. Finally, girls with CAH perform more like males on the AQ (Knickmeyer, Wheelwright, Fane, Hines, & Baron-Cohen, in preparation). This further suggests a role for fT in these areas. We predicted that fT would be negatively correlated to quality of social relationships and positively correlated to restricted interests.

Methods

Participants

Participants were 58 children (35 male, 23 female), age 4.0 to 4.25 years, taking part in a long-term study on the effects of fT. Their mothers had undergone amniocentesis in the Cambridge region between June 1996 and June 1997 and had given birth to the healthy singleton infants between December 1996 and December 1997. All the children had normal chromosomes by cytogenetic analysis. We chose to study these children at age 4 because children's social development has been well studied at this age (Bartsch & Wellman, 1995; Flavell, Miller, & Miller, 1993; Maccoby, 1990).

Outcome variable: the Children's Communication Checklist (CCC) (Bishop, 1998)

The Children's Communication Checklist (CCC) consists of 9 subscales. Scales A (speech) and B (syntax) measure traditional language skills. Scales C to G (inappropriate initiation, coherence, stereotyped conversation, use of context, and rapport) are combined to make a single pragmatic language score. Scale H measures the quality of social relationships. Scale I measures restricted interests. Table 1 shows the possible range of scores on each scale, scores indicating possible impairment and sample items. Note that the scoring is designed so that higher scores always correspond to 'better' performance. At present, reliability and validity for the CCC are restricted to 2 clinical studies. The first study looked at children aged 7 to 9, who have been identified as having language problems (Bishop, 1998). The second included children age 5 to 17 who had been diagnosed with either an autism spectrum condition, attention-deficit hyperactivity disorder, or specific language delay (Bishop & Baird, 2001). A small normally developing comparison group (31 children aged 6 to 16) was also included in the latter study. Alpha indices of internal consistency suggested that the items in each scale were homogenous. Parent–teacher inter-rater reliability ranged from .30 to .64 (Pearson's correlations), all of which were significant at the .01 level.

We chose the CCC for several reasons. Firstly, we were interested in looking at the pragmatics of language. This is an area where sex-differences have been reported (Maccoby, 1999), but the development of pragmatics in childhood has been relatively little studied. The CCC is one of the first instruments designed to explore pragmatics in childhood. Although its reliability and validity are so far restricted, the questions are applicable to anyone using complete sentences. Secondly, the CCC also covers 'traditional' language skills, quality of social relationships, and restricted interests, all of which we were interested in. We could have used a different test for each area, but using a single instrument facilitates the comparison of results. Finally, the CCC was designed, in part, to distinguish children with semantic-pragmatic disorder from children with specific language delay or autism. As discussed earlier, these conditions may represent extreme examples of typical sex-differences. If scores on the checklist are related to fT it could have implications for understanding the sex ratio in these conditions.

Predictor variables

Foetal testosterone levels (fT). The predictor of greatest interest in this study is foetal testosterone. Testosterone 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 (Lutchmaya et al., 2002a, 2002b) (see appendix for details). There were significant differences between male and female testosterone levels (CI −.78 to −.47, $d = 2.0$) (see Table 2). Note, CI is confidence interval. A confidence interval for the difference between two means specifies a range of values within which the difference between the means of the two populations may lie. If the confidence

Table 1 Score ranges and sample items from the Children's Communication Checklist (CCC)

Subscales	Range	Impairment	Sample item
(A) Speech	16–38	<27	1. people can understand virtually everything he/she says
(B) Syntax	24–32	<29	12. speech is mostly 2- or 3-word phrases such as 'me got ball' or 'give dolly'
Pragmatic composite	86–162	<132	
(C) Inappropriate initiation	18–30	<24	16. talks to anyone and everyone
(D) Coherence	20–36	<22	22. it is sometimes hard to make sense of what he/she is saying because it seems illogical or disconnected
(E) Stereotyped conversation	14–30	<24	33. often turns the conversation to a favourite theme, rather than following what the other person wants to talk about
(F) Use of context	16–32	<24	38. tends to repeat back what others have just said
(G) Rapport	18–34	<26	50. makes good use of gestures to get meaning across
(H) Quality of social relationships	14–34	<24	54. is popular with other children
(I) Restricted interests	20–34	<28	65. has one or more overriding specific interests (e.g., computers, dinosaurs) and will prefer doing activities involving this to anything else.

Note: Higher scores correspond to 'better' performance.

Table 2 Means, standard deviations and ranges for outcome and predictor variables by sex

	Girls $N = 23$			Boys $N = 35$			
Variable	Mean	SD	Range	Mean	SD	Range	Cohen's *d*
Age	4.10	.10	4.0–4.25	4.12	.08	4.0–4.25	.22
fT** (nmol/l)	.40	.19	.17–.80	1.04	.40	.13–1.80	2.0
Gestational age at amnio	16.61	1.11	14–19	16.81	1.55	14–21	.19
AFP (MU/l)	10.91	2.82	6.17–19.70	10.42	3.35	3.10–17.90	.21
Oestrogen (pmol/l)	1009	424.63	496–1950	932.55	432.88	404–2630	.18
Maternal age	34.36	4.70	23–40	35.47	4.61	25–43	.23
Paternal age	36.80	6.91	25–53	36.96	5.79	29–52	.03
Maternal education	3.13	.81	2–4	3.27	.83	2–4	.17
Number of older siblings	1.09	.92	0–3	1.10	1.01	0–3	.01
Speech	33.43	3.87	22–38	32.84	3.80	21–37	.15
Syntax	31.10	1.04	28–32	31.30	1.15	28–32	.18
Pragmatic composite	145.28	7.95	130–158	145.97	6.41	131–157	.10
Quality of social relationships	33.00	1.00	31–34	32.38	1.62	29–34	.47
Restricted interests*	32.09	1.60	29–35	30.73	2.30	25–34	.64

**Difference is significant at the .01 level (2-tailed).
*Difference is significant at the .05 level (2-tailed).
Note: *d* is an effect size index. It is calculated by dividing the difference in means for the two groups by the standard deviation. It provides a standardised measure of the magnitude of group differences that can be compared across samples of different size. A *d* of .2 to .4 is considered a small effect size. A *d* between .5 and .7 is considered a medium effect size. A *d* greater than .8 is considered a large effect size (Cohen, 1988).

interval includes 0 it suggests there is no significant difference between the means of the two populations, at a given level of confidence (in this case 95%). The width of the confidence interval gives us some idea about how uncertain we are about the difference in the means. *d* is an effect size index. It is calculated by dividing the difference in means for the two groups by the standard deviation. It provides a standardised measure of the magnitude of group differences that can be compared across samples of different size. A *d* of .2 is considered a small effect size. A *d* of .5 is considered a medium effect size. A *d* greater than .8 is considered a large effect size (Cohen, 1988). We also included the following control variables in our analysis.

Gestational age at amniocentesis. Levels of fT vary during gestation. Although amniocentesis occurs around a specific time-point, this can range from the 14th to the 22nd week. Therefore it was important to determine whether fT was related to gestational age in our sample. Gestational age at amniocentesis was obtained from hospital records. Males showed no linear relationship ($r = .18$, $p < .38$) and no quadratic relationship was apparent. However, for females a significant linear relationship was seen. The correlation between amniotic testosterone and gestational age in our sample was significant ($r = -.54$, $p < .01$). This is unexpected given that Reyes et al. (1974) report no change in foetal serum concentrations of testosterone for females during this period. Examination shows that when the two girls with the highest fT levels are removed the correlation with gestational age is no longer significant. We did not exclude these girls from further analysis because they were within 3 standard deviations of the mean score for fT and gestational age and we wished to keep the sample size as large as possible. They were not indicated as multivariate outliers. To correct for the influence of gestational age we included it as a predictor in the analysis of the combined sexes and when analysing the girls.

Prenatal oestrogen levels. Oestradiol is the most biologically active oestrogen and may have both feminising and masculinising effects on development. It is synthesised in vivo via aromatisation of testosterone and related precursors, so it is important to consider oestradiol when looking at the biological activity of fT. Amniotic oestradiol levels were also assayed by the Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge (see appendix for details). There were no significant differences between oestrogen levels in males and females (CI = −.06–.12, $d = .18$) (see Table 2). Oestrogen levels were positively skewed so a logarithmic transformation was carried out.

Prenatal alpha-foetoprotein level. Alpha-foetoprotein (AFP) is thought to be a general marker for severe foetal ill-health and also provides a specific control for any unexpected abnormalities of amniotic fluid dilution (Wathen, Campbell, Kitau, & Chard, 1993).

Sociodemographic variables. A range of sociodemographic variables were also included in this study because of their possible importance in determining the child's environment. Maternal age, maternal education level, paternal age and number of older siblings could influence the amount and nature of interaction between children, their parents, and their peers. Young children's understanding of beliefs and feelings is influenced by their interactions with their mother and siblings (Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991). Maternal education level was measured according to a 5-point scale: 1 = no formal qualifications, 2 = 'O' level/ GCSE or equivalent, 3 = 'A' level, HND or vocational qualification, 4 = university degree, 5 = postgraduate qualification. Maternal age was particularly important to include because women undergoing amniocentesis have a higher mean age than the general childbearing population. If a particular variable was related to maternal age within our group it would have implications for the wider applicability of our findings.

Results

Descriptive statistics

The first set of analyses provided basic descriptive statistics. Examination of the univariate distributions indicated that several variables had distributions that deviated from the Gaussian distribution. Prenatal oestrogen level was positively skewed (skewness greater than 1) so a natural logarithmic transformation was carried out. The speech, syntax, and restricted interests scales of the CCC were negatively skewed (skewness less than −1). Scores were reflected and then a natural logarithmic transformation was carried out. We investigated the distribution of scores on these CCC subscales to determine whether this was due to a ceiling effect. For the majority of scales there was no significant ceiling effect. For the speech scale, only 4% of the children obtained the maximum score. For the restricted interests scale, only 9% obtained the maximum score. The syntax scale did, however, show some ceiling effects: 57% of children achieved the maximum score. Although it was not transformed, the quality of social relationships scale did show some ceiling effects, with 30% of the children achieving the maximum score. The majority of the children did not achieve a maximum score. Transformation of the scale reduced skewness but increased kurtosis. Residual plots for the regression analyses were better when the untransformed variable was used.

Table 2 presents means, standard deviations, and ranges for each sex separately. Sex-differences were tested for using a *t*-test and equal variances were not assumed. Because several scales of the CCC correlate with each other, both in our sample (see Table 3) and other reported samples, one would usually perform a correction for multiple comparisons. However, because of the sample size we did not have the power to perform these. Therefore we have reported the *t*-test results in terms of confidence intervals instead of *p* values. This provides greater information on the statistical relationships observed. Males had higher fT levels (CI = −.78 to −.47, $d = 2.0$) and more restricted interests (CI = −.21 to −.0154, $d = .64$). There was a trend for females to score better on quality of social relationships (CI = −.13 to 1.32, $d = .47$). The existence of sex-differences on these scales indicates a possible role for fT. Therefore these scales were explored further. Speech, syntax, and pragmatics scores did not show sex-differences, based on confidence intervals and effect sizes. Confidence intervals that do not include 0 are generally considered to be significant.

Examining the correlations between subscales, it is notable that within the group as a whole, quality of social relationships and restricted interests were correlated ($r = .40$, $p < .01$). They were also correlated within the boys ($r = .44$, $p < .01$). These scales are also significantly correlated in the mixed-sex sample reported by Bishop (1998) ($r = .45$ rater A; $r = .65$ rater B). This raises the possibility that scores on both scales are related to single factor.

Relations between outcome variables and foetal testosterone levels

Previous studies investigating fT measured in amniotic fluid have found quadratic relationships with some cognitive measures (Grimshaw et al., 1995a; Lutchmaya et al., 2002b). Examination of scatterplots suggested linear relationships as opposed to quadratic ones for our measures, so a hierarchical multiple regression analysis was used. In the first stage, any predictor variable that correlated significantly with the outcome variable at $p < .2$ was entered into the model (as recommended by Altman, 1991). Suppressor variables were also included when possible; these were predictors that correlated highly ($p < .01$) with the other predictors in the model, but were not significantly correlated with the outcome variable (see Table 4 for correlations between all predictor variables). In the second stage, the main effects of fT and child's sex were tested for inclusion with a stepwise analysis. In the third stage, the interaction of sex and fT was tested for inclusion with a stepwise analysis. This would test whether boys and girls showed consistent differences in terms of the relations between fT and the CCC subscales. Entry criterion was $p < .05$; removal criterion was $p > .1$.

For quality of social relationships the following predictors were entered: gestational age, paternal age, and maternal age (suppressor). Inclusion of fT in the 2nd stage produced a significant F change (F change = 7.071, $p = .012$). A main effect of fT

Table 3 Correlation matrix showing relationships between the Children's Communication Checklist (CCC) scales ($N = 47$–54)

		Speech	Syntax	Quality of social relationships	Restricted interests	Pragmatic composite
Speech	Pearson correlation					
Syntax	Pearson correlation	.64**				
Quality of social relationships	Pearson correlation	−.03	−.15			
Restricted interests	Pearson correlation	−.00	−.19	.40**		
Pragmatic composite	Pearson correlation	.23	−.31*	.25	.25	

**Correlation is significant at the .01 level (2-tailed).
*Correlation is significant at the .05 level (2-tailed).

Table 4 Correlation matrix showing relationships between the independent variables for all subjects of both sexes ($N = 37–58$)

		fT nmol/litre	Gestational age at amnio	AFP (MU/l)	Oestrogen	Maternal age	Paternal age	Maternal education	Number of siblings
fT (nmol/l)	Pearson correlation								
Gestational age at amnio	Pearson correlation	.09							
AFP (MU/l)	Pearson correlation	.03	−.59**						
Oestrogen	Pearson Correlation	.15	−.07	.36**					
Maternal age	Pearson Correlation	.07	−.36*	.17	−.05				
Paternal age	Pearson correlation	−.07	−.37*	.29	.04	.71**			
maternal education	Pearson correlation	−.08	−.26	.12	.04	.24	.29		
Number of siblings	Pearson correlation	.01	−.12	.16	−.16	.37**	.20	.19	

**Correlation is significant at the .01 level (2-tailed).
*Correlation is significant at the .05 level (2-tailed).

was included in the model, while child's sex was excluded. This indicates that fT explains more of the variance than child's sex. This suggests that the sex-differences seen in the scores are testosterone dependent. The interaction of fT and child's sex was also excluded from the model. This suggests that the relationship between fT and quality of social relationships is the same for boys and girls. Those with lower fT levels score higher on quality of social relationships. The only significant predictor in the final model was fT (see Table 5). Residual analysis showed acceptable plots and no outliers.

For restricted interests the following predictors were entered: gestational age, oestrogen and AFP (suppressor). Inclusion of fT in the 2nd stage produced a significant *F* change (*F* change = 6.265, $p = .017$). A main effect of fT was included in the model, while child's sex was excluded. This indicates that fT explains more of the variance than child's sex. This suggests that the sex-differences seen in the scores are testosterone dependent. The interaction of fT and child's sex was also excluded from the model. This suggests that the relationship between fT and restricted interests is the same for boys and girls. Those with higher fT levels had more restricted interests. The only significant predictor in the final model was fT, although gestational age also approached significance (see Table 5). Residual analysis showed acceptable plots and no outliers.

The analyses did not indicate different relationships with fT and quality of social relationships or restricted interests for boys and girls. Also a main effect of sex was excluded as a predictor, but fT was included. This suggests that the sex-differences seen in the scores are testosterone dependent. However, to further investigate whether the previous result might be due to a sex-difference (not necessarily involving testosterone), we analysed the relationship between these scores and fT within each sex. It should be kept in mind that this reduced the sample-size by half and therefore the power of the analysis.

For quality of social relationships no significant relationship with fT was observed for boys or girls.

Within boys none of the background variables correlated significantly with restricted interests at $p < .2$. fT was retained in the model following the regression analysis ($\beta = .172$, $p = .052$) and the model explained 12.4% of the variance in restricted interests. Boys with higher fT levels had more restricted interests. Residual analysis showed acceptable plots and no outliers. fT and restricted interest did not correlate in girls ($r = -.21$, $p = .358$), therefore a regression analysis was not performed for girls.

Table 5 Hierarchical regression analysis, final model: sexes combined

Dependent variable	R^2	Predictors	B	SE B	Sig.
Quality of social relationships	.22	Constant	35.674	4.769	.000
		Gestational age at amnio	−.097	.191	.616
		Maternal age	.014	.093	.878
		Paternal age	−.021	.074	.784
		fT	−1.636	.615	.012*
Restricted interests[a]	.25	Constant	.806	.775	.305
		Gestational age at amnio	−.054	.027	.055
		AFP	−.020	.012	.118
		Oestrogen[b]	.265	.196	.185
		fT	.173	.069	.017*

Note: For quality of social relationships: power > .60 For restricted interests: power > .75.
[a]Scores were reflected and logged before analysis. [b]Scores were logged before analysis.
*$p < .05$.

Discussion

This study tested for correlations between foetal testosterone (fT) and the 5 subscales of the Children's Communication Checklist (CCC) (Bishop, 1998). We predicted that no sex-differences would be seen on the first two scales of the CCC, which measure non-social aspects of language skills. In contrast, we predicted that females would score better on pragmatic language abilities as measured by the pragmatic composite and on quality of social relationships and that this would be related to fT. We also predicted that boys would have more restricted interests and that this would be related to fT. In general our predictions were supported. No sex-differences were seen on the first two scales of the CCC. There was a trend for girls to score better on quality of social relationships and this was related to fT levels in the group as a whole. There was a significant difference between boys and girls on the restricted interests scale. Boys had more restricted interests and this was related to fT levels in the group as a whole. fT was also related to restricted interests when boys were examined separately.

Sex-difference and pragmatic language

There was a lack of any significant sex-difference on the pragmatic composite. It is possible that sex-differences in pragmatic language are not present, at least in this age group and on this particular test. However, it is also possible that the analysis was underpowered. The sample size used is comparable to other studies using amniotic fluid, but might not be sufficient to detect differences on this scale. It should be noted that a significant sex-difference for the pragmatic subscale was observed in a sample 6-year-old twins when completed by their teachers (girls: mean (SD) 153.08 (6.9); boys: mean (SD) 146.91 (10.15); $t = 3.92$; $p < .0001$) (Bishop and Laws, personal communication). The difference in our sample may reflect a type II error in our study, but could also indicate an age effect. Sex-differences in pragmatic language may arise after age 4. Finally, the lack of relationship might reflect the fact that the pragmatic composite comprises a mix of cognitive skills.

Quality of social relationships and restricted interests

There was a trend for girls to score better on quality of social relationships and there was a main effect of fT on this scale when the group was examined as a whole (the analysis excluded a main effect of sex, or an interaction of sex and fT). This indicates that in both boys and girls, higher fT levels are associated with poorer quality of social relationships. Causal interpretations are, of course, unjustified from this correlational study, but the observed correlations are consistent with the hypothesis that fT diminishes social cognition or social interest in both boys and girls, resulting in poorer quality of social relationships. Another explanation would be that relations between fT and social relationships are mediated by other variables, or that fT is serving as an index for an unknown third variable. It should be noted that in the sample of 6-year-old twins reported earlier no sex-differences were seen on this scale (Bishop and Laws, personal communication). However, significant sex-differences are seen on similar instruments such as the Social Cognitive Skills Questionnaire (Scourfield et al., 1999; Skuse et al., 1997) and the Social Responsiveness Scale (Constantino, Przybeck, Friesen, & Todd, 2000; Constantino & Todd, 2000). No significant relationships were seen when the sexes were examined separately, but this may be because the analysis was underpowered. A sample size of approximately 80 would be required to give the model a power of .6, assuming a similar effect size as was detected when both sexes were examined together. It would be necessary to run this experiment with a larger sample before drawing any strong conclusions.

Boys had significantly more restricted interests than girls. It should be noted that in the sample of 6-year-old twins reported earlier no sex-differences were seen on this scale (Bishop and Laws, personal communication). However, higher male scores on the imagination and attention scales of the AQ do suggest that males may have more restricted interests than females (Baron-Cohen et al., 2001b). There was a main effect of fT on this scale when the group was examined as a whole (the analysis excluded a main effect of sex, or an interaction of sex and fT). This indicates that in both boys and girls, higher fT levels are associated with more restricted interests. This relationship was also seen when boys were examined separately. It is possible that the sample size was too small to observe a relation in girls (there were 35 boys in the sample, but only 23 girls). A sample size of approximately 200 would be required to give the model a power of .6, assuming a similar effect size as was detected in the boys. It is also possible that because the lower range of fT in this sample is close to the detection limit of the testosterone assay, less variability is detectable in girls than boys. (See Appendix for more detailed discussion of the hormone assay). However, in the study by Finegan et al. (1992), the lowest fT levels in girls were also close to the detection limit, but in that case, fT-behaviour relationships were seen in girls but not boys.

As with the observed relations between fT and social relationships, the relation between fT and restricted interests may indicate that fT increases restricted interests, but it is also possible that relations between fT and restricted interests are mediated by other variables or that fT is serving as an index for an unknown third variable.

The quality of social relationships and restricted interests scales were included in the design of the CCC in order to determine whether a diagnosis of autism should be considered (Bishop, 1998). As mentioned in the introduction, autism has been described as an extreme manifestation of some sexually dimorphic traits (Baron-Cohen, 2002). Our results are compatible with this framework. In our study, normal boys had a poorer quality of social relationships than normal girls. People with autism have even greater difficulty with social relationships. In our study, normal boys had more restricted interests than girls. People with autism have even more restricted interests. Our results suggest that high fT levels are associated with poorer quality of social relationships and more restricted interests, particularly in boys. Although we cannot extrapolate directly from this study to autism, our results suggest that it may be worth while to explore whether fT is involved in the male vulnerability to autism. A study comparing fT levels in children who go on to develop autism with normally developing children, and children referred for other developmental disorders, is currently under way in our lab. It is also possible that testosterone promotes a general 'male vulnerability' (Kraemer, 2000). It is worth keeping in mind that testosterone is not the only factor that varies between males and females. Arnold (1996) suggests that there may be genetically triggered differences between males and females that do not involve hormone intermediates. Skuse et al. (Skuse, 2000; Skuse et al., 1997), for example, suggest that an imprinted gene on the X chromosome is responsible for the sex ratio in autism. It is also possible that there is an interaction between these postulated factors.

Limitations

When evaluating our results, it is important to keep in mind that our sample of 4-year-old, typically developing children is different from those populations for which published data on the reliability and validity of the CCC are available. Comparing the scores for our sample to the 31 normally developing controls reported in Bishop and Baird (2001), mean scores are very similar; the speech and syntax scores are slightly lower; the pragmatic composite about 8 points lower. The speech and syntax scores were also slightly more variable in our sample. We realise that relying on maternal report has some drawbacks, notably that different mothers may interpret items differently. However, the skills we were exploring did not lend themselves to laboratory testing. A crucial part of pragmatic language is the ability to adjust speech for different contexts and speakers. Likewise, social skills will vary widely depending on whether a child is relating to a parent, stranger, or a friend. A single or even a series of laboratory sessions could not take account of this variety. An advantage of maternal report is that mothers have the opportunity to judge their children's skills in a variety of contexts over an extended period of time. In the future, home observations by skilled observers while mothers execute semi-standardised interactive communication tasks with their children could be performed. This would combine ecological validity with less biased observations.

The study is also limited by the problems inherent in studying foetal endocrinology. These include a lack of knowledge about the correlation between amniotic and plasma levels of foetal testosterone and difficulties in determining foetal age at amniocentesis. We have calculated gestational age using the date of last menstrual period. Slightly different estimates might be obtained using sonographic measures (such as femur length). The actual effect of testosterone will not depend solely on the total extractable amount found in amniotic fluid. Binding proteins and degradation enzymes also affect the availability of the hormone. Presence and sensitivity of appropriate receptors determines whether and how potent testosterone's effects may be.

Finally, we have also been restricted to studying a non-random population. It should be noted that previous studies investigating the relationship of fT to cognitive development in humans have relied upon individuals with abnormal hormonal environments during pregnancy such as individuals with congenital adrenal hyperplasia or those exposed to drugs that mimic or block natural hormones. In these cases it is difficult to differentiate between the effects of the hormonal environment and those of any gene abnormality associated with the disorder or any additional effects the drug may produce. 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 (for example, late maternal age) will be shared by all the subjects. It is unlikely that amniotic testosterone levels are different in mothers who undergo amniocentesis compared to those who do not, because within this group no relationship was seen between fT and maternal age, alphafoetoprotein level, paternal age, or parental education level (Lutchmaya, 2000).

Scarr and McCartney (1983) and others have pointed out that there is no such thing as a purely biological or purely environmental contribution to behaviour. Addressing all potential factors and their interactions would require an immense sampling size. Our focused, relatively small-scale study is valuable in demonstrating that foetal testosterone potentially has long-term effects on social relationships and restricted interests. Finally, the study raises the intriguing possibility that quality of social relationships and restricted interests, instead of being two independent dimensions as proposed (Baron-Cohen, 2002), may be controlled by a single factor.

Acknowledgements

Rebecca Knickmeyer was supported by a British Government Overseas Research Studentship (ORS), the Cambridge Overseas Trust and Gonville and Caius College (University of Cambridge). Additional funding was generously donated by the Lurie Marks Foundation and the Medical Research Council. This work was submitted in part fulfilment of the degree of PhD, University of Cambridge. Simon Baron-Cohen was supported by the Medical Research Council (MRC) UK and the Isaac Newton Trust during the period of this work. We are grateful to Ian Goodyer, Gerry Hackett, Svetlana Lutchmaya, and Rosa Hoekstra for their support and discussion. Finally, we are indebted to the families who participated in this study.

Correspondence to

Rebecca Knickmeyer, Autism Research Centre, Department of Experimental Psychology and Psychiatry, University of Cambridge, Douglas House, 18b Trumpington Rd, Cambridge CB2 2AH, UK; Tel: 01223 746030; Fax: 01223 746033; Email: rk250@cam.ac.uk

References

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*, 231–239.

Altman, D. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

Arnold, A. (1996). Genetically triggered sexual differentiation of brain and behavior. *Hormones and Behavior*, *30*, 495–505.

Arnold, A., & Gorski, R. (1984). Gonadal steroid induction of structural sex differences in the central nervous system. *Annual Review of Neuroscience*, *7*, 413–442.

Banerjee, M. (1997). Hidden emotions: Preschoolers' knowledge of appearance-reality and emotion display rules. *Social Cognition*, *15*, 107–132.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*, 248–254.

Baron-Cohen, S. (2003). *The essential difference*. London: Allen Lane, Penguin Books.

Baron-Cohen, S., & Hammer, J. (1997). Is autism an extreme form of the male brain? *Advances in Infancy Research*, *11*, 193–217.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, *38*, 813–822.

Baron-Cohen, S., Knickmeyer, R., & Lutchmaya, S. (2004). *Prenatal testosterone in mind: Amniotic fluid studies*. Cambridge, MA: MIT Press/Bradford Books.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger Syndrome or High Functioning Autism. *Journal of Autism and Developmental Disorders*, *29*, 407–18.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The Systemising Quotient (SQ): An investigation of adults with Asperger Syndrome or High Functioning Autism and normal sex differences. *Philosophical Transactions of the Royal Society, Series B, Special issue on 'Autism: Mind and Brain*, *358*, 361–374.

Baron-Cohen, S., Ring, H., Bullmore, E., Wheelwright, S., Ashwin, C., & Williams, S. (2000). The amygdala theory of autism. *Neuroscience and Behavioural Reviews*, *24*, 355–364.

Baron-Cohen, S., Wheelwright, S., & Hill, J. (2001a). The 'Reading the mind in the eyes' test revised version: A study with normal adults, and adults with Asperger Syndrome or High-Functioning autism. *Journal of Child Psychiatry and Psychiatry*, *42*, 241–252.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001b). The Autism Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High Functioning Autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17.

Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. Oxford: Oxford University Press.

Benenson, J. (1993). Greater preference among females than males for dyadic interaction in early childhood. *Child Development*, *64*, 544–555.

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

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.

Bishop, D. (1990). *Handedness and developmental disorder*. Oxford: Blackwell.

Bishop, D. (1998). Development of the children's communication checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry*, *6*, 879–891.

Bishop, D. (2001). Genetic and environmental risks for specific language impairment in children. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences*, *356*, 369–380.

Bishop, D., & Baird, G. (2001). Parent and teacher report of pragmatic aspects of communication: Use of the Children's Communication Checklist in a clinical setting. *Developmental Medicine and Child Neurology*, *43*, 809–818.

Breedlove, S. (1994). Sexual differentiation of the human nervous system. *Annual Review of Psychology*, *45*, 389–418.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Collaer, M., & Hines, M. (1995). Human behavioural sex differences: A role for gonadal hormones during early development? *Psychological Bulletin*, *118*, 55–107.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Ba'tki, A., & Ahluwalia, J. (2001). Sex differences in human neonatal social perception. *Infant Behavior and Development*, *23*, 113–118.

Constantino, J.N., Przybeck, T., Friesen, D., & Todd, R.D. (2000). Reciprocal social behaviour in children with and without pervasive developmental disorders. *Developmental and Behavioural Pediatrics*, *1*, 2–11.

Constantino, J., & Todd, R. (2000). Genetic structure of reciprocal social behavior. *American Journal of Psychiatry*, *157*, 2043–2045.

De Vries, G., Rissman, E., Simerly, R., Yang, L., Scordalakes, E., Auger, C., Swain, A., Lovell-Badge, R., Burgoyne, P., & Arnold, A. (2002). A model system for study of sex chromosome effects of sexually dimorphic neural and behavioral traits. *The Journal of Neuroscience*, *22*, 9005–9014.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development*, *62*, 1352–1366.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, Serial no. 242, vol. 59*.

Finegan, J., Niccols, G., & Sitarenios, G. (1992). Relations between prenatal testosterone levels and cognitive abilities at 4 years. *Developmental Psychology*, *28*, 1075–1089.

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.

Fitch, R., & Bimonte, H. (2002). Hormones, brain, and behavior: Putatative biological contributions to cognitive sex differences. In A. McGillicuddy-De Lisi & R. De Lisi (Eds.), *Biology, society, and behavior: The development of sex differences in cognition*. Greenwich, CT: Ablex.

Fitch, R., & Denenberg, V. (1998). A role for ovarian hormones in sexual differentiation of the brain. *Behavioral and Brain Sciences*, *21*, 311–352.

Flavell, J., Miller, P., & Miller, S. (1993). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Geary, D. (1998). *Male, female: The evolution of human sex differences*. Washington, DC: American Psychological Association.

Geary, D. (2002). Sexual selection and sex differences in social cognition. In A. McGillicuddy-De Lisi & R. De Lisi (Eds.), *Biology, society, and behavior: The development of sex differences in cognition*. Greenwich, CT: Ablex.

Geschwind, N., & Galaburda, A.M. (1985). Cerebral Lateralisation: Biological mechanisms, associations and pathology: III. A hypothesis and a program for research. *Archives of Neurology*, *42*, 634–654.

Gitter, A., Black, H., & Mostofsky, B. (1972). Race and sex in the perception of emotion. *Journal of Social Issues*, *28*, 63–78.

Gleason, J., & Ely, R. (2002). Gender differences in language development. In A. McGillicuddy-De Lisi & R. De Lisi (Eds.), *Biology, society and behavior: The development of sex differences in cognition*. Greenwich, CT: Ablex Publishing.

Goy, R., & McEwen, B. (1980). *Sexual differentiation of the brain*. Cambridge, MA: The MIT Press.

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

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

Hall, J.A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, *85*, 845–858.

Hall, J.A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: John Hopkins University Press.

Halpern, D. (1992). *Sex differences in cognitive abilities*. Hillsdale, NJ: Erlbaum.

Hampson, E., Rovet, J., & Altmann, D. (1998). Spatial reasoning in children with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Developmental Neuropsychology*, *14*, 299–320.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, *66*, 843–855.

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

Hines, M., Golombok, S., Rust, J., Johnston, K., Golding, J., & Team t. A. L. S. o. P. a. C. S. (2002). Testosterone during pregnancy and gender role behavior of preschool children: A longitudinal, population study. *Child Development*, *73*, 1678–1687.

Hines, M., & Sandberg, E. (1996). Sexual differentiation of cognitive abilities in women exposed to diethylstilbestrol (DES) prenatally. *Hormones and Behavior*, *30*, 354–363.

Hines, M., & Shipley, C. (1984). Prenatal exposure to diethylstilbestrol (DES) and the development of sexually dimorphic cognitive abilities and cerebral lateralization. *Developmental Psychology*, *20*, 81–94.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early language growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–248.

Hyde, J.S., & Linn, M.C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*, 53–69.

Jacklin, C.N., Maccoby, E.E., & Doering, C.H. (1983). Neonatal sex-steroid hormones and timidity in 6–18-month-old boys and girls. *Developmental Psychobiology*, *16*, 163–168.

Jacklin, C.N., Wilcox, K.T., & Maccoby, E.E. (1988). Neonatal sex-steroid hormones and cognitive abilities at six years. *Developmental Psychobiology*, *21*, 567–574.

Jolliffe, T., & Baron-Cohen, S. (1997). Are people with autism or Asperger's syndrome faster than normal on the Embedded Figures Task? *Journal of Child Psychology and Psychiatry*, *38*, 527–534.

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: The MIT Press.

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

Kraemer, S. (2000). The fragile male. *British Medical Journal*, *321*, 1609–1612.
Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systemising in adults with and without Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *34*, 301–310.
Lutchmaya, S. (2000). *Foetal testosterone and social-communicative development*. Unpublished PhD 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.
Maccoby, E.E. (1990). Gender and relationships: A developmental account. *American Psychologist*, *45*, 513–520.
Maccoby, E.E. (1999). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.
MacLusky, N., & Naftolin, F. (1981). Sexual differentiation of the central nervous system. *Science*, *211*, 1294–1303.
Martin, C.R. (1985). *Endocrine physiology*. New York: Oxford University Press.
Meyer-Bahlburg, H., Ehrnhardt, A., Rosen, L., Gruen, R., Veridiano, N., Vann, F., & Neuwalder, H. (1995). Prenatal estrogens and the development of homosexual orientation. *Developmental Psychology*, *31*, 12–21.
Miller, E. (1994). Prenatal sex hormone transfer: A reason to study opposite-sex twins. *Personality and Individual Differences*, *17*, 511–529.
Nagami, M., McDonough, P., Ellegood, J., & Mahesh, V. (1979). Maternal and amniotic fluid steroids throughout human pregnancy. *American Journal of Obstetrics and Gynecology*, *134*, 674–680.
Nass, R., Baker, S., Speiser, P., Virdis, R., Balsamo, A., Cacciari, E., Loche, A., Dumic, M., & New, M. (1987). Hormones and handedness: Left-hand bias in female adrenal hyperplasia patients. *Neurology*, *37*, 711–715.
Plante, E., Boliek, C., Binkiewicz, A., & Erly, W. (1996). Elevated androgen, brain development and language/learning disabilities in children with congenital adrenal hyperplasia. *Developmental Medicine and Child Neurology*, *38*, 423–37.
Reinisch, J., & Sanders, S. (1992). Effects of prenatal exposure to diethylstilbestrol (DES) on hemispheric laterality and spatial ability in human males. *Hormones and Behavior*, *26*, 62–75.
Resnick, S., Gottesman, 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 gonadotropin and sex steroid concentrations. *Journal of Clinical Endocrinology and Metabolism*, *38*, 612–617.
Reynell, J., & Huntley, M. (1985). *Reynell development language scales* (2nd rev.). Windsor, England: NFER-Nelson.
Robinson, J., Judd, H., Young, P., Jones, D., & Yen, S. (1977). Amniotic fluid androgens and estrogens in midgestation. *Journal of Clinical Endocrinology*, *45*, 755–761.
Rodgers, C., Fagot, B., & Winebarger, A. (1998). Gender-typed toy play in dizygotic twin pairs: A test of hormone transfer theory. *Sex Roles*, *39*, 173–184.
Rotter, N.G., & Rotter, G.S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior*, *12*, 139–148.
Rutter, M. (1978). Language disorder and infantile autism. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment*. New York: Plenum.
Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype->environment effects. *Child Development*, *54*, 424–435.
Schacter, S. (1994). Handedness in women with intrauterine exposure to diethylstilbestrol. *Neuropsychologia*, *32*, 619–623.
Scourfield, J., Martin, N., Lewis, G., & McGuffin, P. (1999). Heritability of social-cognitive skills in children and adolescents. *British Journal of Psychiatry*, *175*, 559–564.
Skuse, D.H. (2000). Imprinting, the X-chromosome, and the male brain: Explaining sex differences in the liability to autism. *Pediatric Research*, *47*, 9–16.
Skuse, D., James, R., Bishop, D., Coppins, B., Dalton, P., Aamondt-Leeper, G., Bacarese-Hamilton, M., Creswell, C., McGurk, R., & Jacobs, P. (1997). Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature*, *387*, 705–708.
Smith, L., & Hines, M. (2000). Language lateralization and handedness in women prenatally exposed to diethylstilbestrol (DES). *Psychoneuroendocrinology*, *25*, 497–512.
Stodgell, C., Ingram, J., & Hyman, S. (2001). The role of candidate genes in unraveling the genetics of autism. *International Review of Research in Mental Retardation*, *23*, 57–81.
Tierney, I., Smith, L., Axworthy, D., & Ratcliffe, S. (1984). The McCarthy scales of children's abilities-sex and handedness effects in 128 Scottish five-year-olds. *British Journal of Educational Psychology*, *54*, 101–105.
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.
Williams, C., & Meck, W. (1991). The organizational effects of gonadal steroids on sexually dimorphic spatial ability. *Psychoneuroendocrinology*, *16*, 155–176.
Wing, L. (1981). Asperger Syndrome: A clinical account. *Psychological Medicine*, *11*, 115–130.
Zahn-Waxler, C., Robinson, J., & Emde, R. (1992). The development of empathy in twins. *Developmental Psychology*, *28*, 1038–1047.

Manuscript accepted 2 February 2004

Appendix: Hormone assays

Testosterone

Amniotic fluid was extracted with diethyl ether. Recovery experiments have demonstrated consistent 95% 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 labelled testosterone analogue. The detection limit of the assay using the ether-extraction method is approximately .1 nmol/l. This method measures total extractable testosterone. Given that those with the lowest fT levels in our sample were near this limit, there was a 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 fT. Only 2 girls (about 9% of the female sample) scored below .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 (although transformation of scores was not necessary). This could be one reason why more relations were seen in boys than girls.

Oestrogen

Amniotic fluid was extracted with diethyl ether. Recovery experiments have demonstrated consistent 95% recovery of oestradiol using this method. The estradiol was measured by fluorescence-labelled 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-labelled 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 foeto-placental unit, cross react to less than .9% by weight. Intra-assay coefficients of variation (i.e., 1 standard deviation expressed as a percentage of the mean value) were 5.2% at 180 pmol/L and 3.9% at 875 pmol/L.