Available online at www.sciencedirect.com
SCIENCE DIRECT®
Hormones and Behavior 49 (2006) 282 – 292
Hormones and Behavior
www.elsevier.com/locate/yhbeh

Fetal testosterone and empathy

Rebecca Knickmeyer [a,*], Simon Baron-Cohen [a], Peter Raggatt [b], Kevin Taylor [b], Gerald Hackett [c]

[a] Autism Research Centre, Department of Psychiatry, University of Cambridge, Douglas House, 18b Trumpington Rd, Cambridge CB2 2AH, UK
[b] Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK
[c] Rosie Maternity Hospital, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK

Received 31 March 2005; revised 8 July 2005; accepted 23 August 2005
Available online 13 October 2005

Abstract

Background: In animals, fetal testosterone (fT) plays a central role in organizing the brain and in later social behavior. In humans, exposure to atypical levels of prenatal androgens may result in masculine behavior and ability patterns. Normal inter-individual variation in fT levels has also been correlated with later sex-typed behavior.
Methods: In the current study, 38 children (24 male, 14 female), whose fT was analyzed in amniotic fluid, were followed up at age 4. They were asked to describe cartoons with 2 moving triangles whose interactions with each other suggested social relationships and psychological motivations.
Results: Females used more mental and affective state terms to describe the cartoons than males. fT was not associated with the frequency of mental or affective state terms. Females also used more intentional propositions than males. fT was negatively correlated with the frequency of intentional propositions, taking sex differences into account. fT was also negatively correlated with the frequency of intentional propositions when males were examined separately. Males used more neutral propositions than females. fT was directly correlated with the frequency of neutral propositions, taking sex differences into account. This relationship was not seen when males and females were examined separately.
Conclusions: These findings implicate fT in human social development. The relevance of our findings to the 'extreme male brain' theory of autism is also discussed.

Keywords: Fetal testosterone; Social development; Empathy; Theory of mind; Autism

Experiments in non-human mammals show that fetal testosterone (fT) plays a critical role in a wide range of sex differences, although the contribution of estrogen of ovarian origin (Fitch and Bimonte, 2002; Fitch and Denenberg, 1998) and direct genetic influences on sexual differentiation (Arnold, 1996; De Vries et al., 2002) are increasingly recognized. fT has been shown to affect the anatomy of the brain, including the hypothalamus, limbic system, and neocortex (Arnold and Gorski, 1984; Breedlove, 1994; MacLusky and Naftolin, 1981), sexually dimorphic behaviors such as aggression and activity level (Goy and McEwen, 1980), and sexually dimorphic cognitive abilities such as spatial navigation (Williams and Meck, 1991).

In human beings, sex differences are apparent both in brain structures and cognitive skills (Breedlove, 1994; Collaer and Hines, 1995; Halpern, 1992; Kimura, 1999; MacLusky and Naftolin, 1981). 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 cognition. 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 'empathizing.' This is defined 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.

Both authors refer to evolutionary arguments to explain the female advantage in empathy or social cognition: First, because females historically migrated to the social group of their mate (while males remained in their birth group), females have been forced to form social alliances with non-kin (Geary, 1998). Arguably, being accepted into a new social group requires better social skills than staying within one's kin group. Secondly,

* Corresponding author. Fax: +44 1223 746033.
E-mail address: rk250@cam.ac.uk (R. Knickmeyer).

0018-506X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.yhbeh.2005.08.010

women's role as primary care-giver for children may also have created selection pressures for greater empathy in order to read her infant's mental states and needs rapidly, thus promoting the infant's survival. Thirdly, women's role in child care would have been significantly helped by the ability to form close relationships with other women to obtain social support. Fourthly, concern for others could have been disadvantageous for males as their reproductive success may have depended on dominating others, sometimes through physical aggression (Baron-Cohen, 2003; Geary, 1998). Finally, when females do compete with each other, they use indirect methods such as gossip and social exclusion (Crick et al., 1997) in an attempt to damage the social networks crucial for female success and reduce their competitors' desirability as mates. Although evolutionary explanations are not directly testable, these arguments give rise to the notion that the female advantage in social skills may not be solely due to cultural factors but may also be in part biological.

Sex differences are also seen in neurodevelopmental conditions involving social and communicative development. Thus, 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 form of some sexually dimorphic traits or an extreme of the male brain (Baron-Cohen, 2002). Autism spectrum conditions (ASCs) are characterized by impairments in reciprocal social interaction, verbal and non-verbal communication, and imaginative play. In addition, the individual shows unusually strong repetitive behavior and narrow interests (APA, 1994; ICD-10, 1994). As many as 1 in every 200 people may have an ASC (Scott et al., 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 et al., 2001), but perform better than people without autism on tests where males usually outperform females, such as the "Embedded Figures Task" (Joliffe and Baron-Cohen, 1997; Baron-Cohen and Hammer, 1997). It has been suggested that increased levels of fT may produce excessive masculinization of the brain and thereby increase the risk for autism spectrum conditions (Baron-Cohen et al., 2004).

fT can be measured in amniotic fluid collected during mid-trimester amniocentesis (Finegan et al., 1989). fT is thought to enter the amniotic fluid via diffusion through the skin in early pregnancy and later from fetal urination (Klopper, 1970; Nagami et al., 1979; Robinson et al., 1977). Although the exact correlation between fT levels in the serum and the amniotic fluid is unknown, the maximal sex difference in amniotic fT occurs between weeks 12 and 18, closely paralleling peak serum levels (Finegan et al., 1989). 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 and Hines, 2000). Therefore, it is likely that this is an important period for sexual differentiation of the brain in humans as well. Using this paradigm, we have shown that fT is negatively related to vocabulary size at 12 months of age (Lutchmaya et al., 2002b), amount of eye contact at 12 months of age (Lutchmaya et al., 2002a), and quality of social relationships at age 4 (Knickmeyer et al., 2005). The results suggest that exposure to high levels of fT may result in poorer social cognition.

The above study of social relationships and fT (Knickmeyer et al., 2005) relied on maternal report, using the Children's Communication Checklist (CCC) (Bishop, 1998). While questionnaires have the advantage of rapid data collection, this approach has some drawbacks, notably that different mothers may interpret items differently. To further test the apparent relationship between fT and social development, we therefore invited these children and their families for a laboratory test of social cognition, reported below.

Research into social cognition in childhood has primarily focused on theory of mind (ToM), the ability to attribute independent mental states to others and to oneself in order to explain and predict behavior. ToM also constitutes an important component of "empathizing" (Baron-Cohen, 2002, 2003). The gold-standard laboratory test for ToM has been the false belief task, where a participant needs to be aware of a character's mistaken mental state (as opposed to the participant's own belief) in order to predict that character's behavior. False belief tasks have played an important role in tracking the development of ToM in young children and in characterizing the difficulties in social understanding typical of autism. However, these tests are usually dichotomous in nature (the child either passes or fails) and so cannot capture the continuum of social ability (or disability) displayed by children. They also have a limited power to detect subtle individual differences, including sex differences (Baron-Cohen et al., 1997), and are therefore not useful in a study relating differences to variations in fT. Therefore, we chose to use a different type of task: one that explored children's capacity to spontaneously attribute social qualities to ambiguous visual stimuli.

This paradigm was first created by Heider and Simmel (1944). They showed a film of two triangles and a circle moving within and around a rectangle and found that adults viewed the moving shapes as having goals and intentions and that they described the moving shapes with agentive and mental state terms. Springer et al. (1996) showed Heider and Simmel's film to 3-, 4-, and 5-year-old children and found that character attributions were more differentiated in older than younger children. Montogomery and Montogomery (1999) showed that even 3-year-olds infer an intention or goal of an animated shape, on the basis of its simple pattern of motion. Bowler and Thommen (2000) showed the original Heider and Simmel animation to children with autism or Asperger syndrome and controls. They found that the groups were equally able to distinguish biological (agentive) from mechanical motion and used comparable amounts of mental state language when viewing actions between animate agents (such language was infrequent in all groups). Abell et al. (2000) suggested that, although the original Heider and Simmel animation does involve goal-directed action, it may not easily elicit mental state attribution, and this may explain the lack of differences between groups in the Bowler and Thommen study.

More recently, two independent groups have designed animation sequences specifically designed to elicit mental state attributions. Klin (2000) found that individuals with autism showed marked deficits on all seven indices of social cognition on his Social Attribution Task (SAT), including the number of mental and affective state terms used. Abell et al. (2000) found that, although children with autism used mentalizing descriptions less often than normally developing 8-year-olds when describing their computer-presented animations, they did so as often as children with general intellectual impairment. However, children with autism frequently referred to mental states that were inappropriate to the animation. No one has examined whether there is a sex difference in the tendency to attribute social meaning to these films, but research on sex differences in social cognition suggests that a difference (female advantage) may exist.

We showed a series of computer-presented animations to children whose fT levels had been measured in amniotic fluid. We predicted that girls would use more intentional language than boys and that these differences would be related to inter-individual variation in fT.

Methods

Participants

Participants were $n = 39$ children (25 male, 14 female), age 4.0 to 4.25 years, taking part in a long-term study on the effects of fT. These children represent a subset of the 58 children whose mothers completed the Children's Communication Checklist in our earlier study (Knickmeyer et al., 2005). Many of the families in the sample live several hours from our testing center and given the work schedules of the parents and the fact that all children are enrolled in various schools, traveling for testing is difficult and time-consuming. This meant that not all families were available for laboratory tests.

Outcome variable

Computer-presented animations were provided by Fulvia Castelli and were used in the studies by Abell et al. (2000) and Castelli et al. (2000). The animations showed one large red and one small blue triangle moving around a screen which contained a rectangular enclosure. One animation (random) showed the triangles moving about purposelessly (bouncing off the sides) and not interacting with each other (this was chosen from a possible set of 4). Two animations were designed to convey ToM (these were taken from a possible set of 6). One film showed the big triangle coaxing the little one out of the enclosure (see Fig. 1). One showed the little triangle hiding behind a door and surprising the big triangle. Sequences lasted 34 to 45 s each. Children watched the film once and were then asked what was happening. They were then asked to describe the film again as it was playing (to reduce the memory burden). For the random film, only these initial descriptions were recorded. For the ToM films, after the first descriptions were given, children were asked a series of questions designed to elicit more information and to encourage them to view the sequence in human terms. Fig. 2 shows a sample transcript with the interviewer's questions and the child's response. The children's narratives were tape-recorded and then transcribed.

A rater, blind to the children identities, counted the number of mental state terms and affective state terms used. Following Bartsch and Wellman's (1995) taxonomy, mental state terms included terms expressing one character's belief, desire, thought, imagination, intention, plan, motivation, and behaviors that could not exist without shared cognition between characters (e.g., tricking, bullying, sneaking, hiding). Affective state terms included emotional terms (e.g., happy, sad, afraid) and behaviors that could not exist without a shared emotional state between the characters (cheering, hugging, kissing). Behaviors which were not uniquely human and/or simply implied an emotion were not included (e.g., fighting, playing).

In order to control for the length of narrative, the number of terms was divided by the total number of propositions used; a proposition is defined as a verb plus its complement. 22 of 38 transcripts were rated by a second rater. Inter-rater reliability for mental state terms was 0.88. Inter-rater reliability for affective state terms was 0.68. Inter-rater reliability for number of propositions was 0.98. We predicted that females would use more mental and affective state terms than males.

Each proposition was also defined as one of the types set out by Bowler and Thommen (2000). The resulting classes of events are undirected actions (Act), action between animates (ActA), action on the rectangle (ActR), relation with the rectangle (RelR), intentional (Int), and neutral (Neu). ActA and ActR can be considered a subset of 'intentional acts.' The category intentional refers to all intentional actions that do not fall into ActA or ActR categories. In practice, this contained many propositions describing emotional states, beliefs, and desires. Bowler and Thommen (2000) found that individuals with autism made less reference to actions between animate objects (ActA) than controls. We therefore predicted that females would make more ActA propositions than males. We also predicted that females would make more Int propositions than males, if they are more likely to ascribe intentional actions to the stimuli in general. Table 1 provides examples of each proposition type. Number of each type of proposition was divided by the total number or propositions to control for length of narrative. Inter-rater reliability for each category is also presented in Table 1. Inter-rater reliability was good to excellent, except in the case of relation to rectangle (RelR). This may reflect the difficulty in ascertaining whether the children were referring to the rectangular enclosure or the screen itself when describing the movements of the triangles.

Predictor variables

Fetal testosterone levels (fT) (nmol/l)

The predictor of greatest interest in this study is fetal testosterone. T levels in amniotic fluid were measured by radioimmunoassay by the

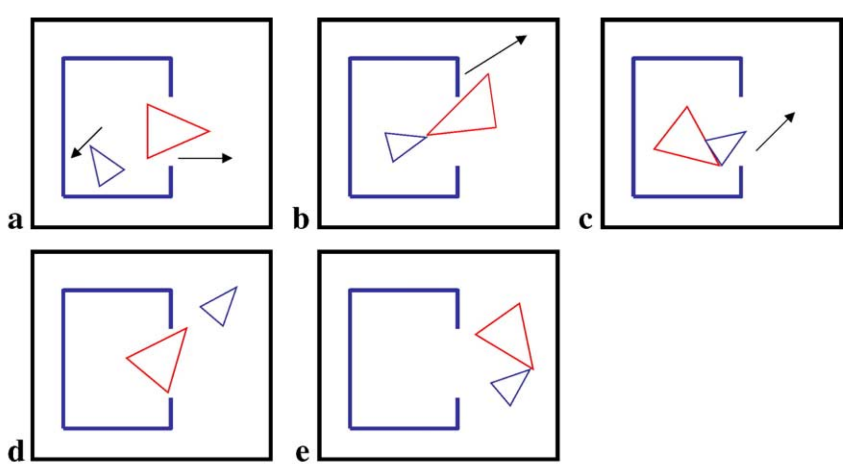

Fig. 1. Stills from animation scripted as 'Coaxing'. (a) Mother tries to interest child in going outside. (b) Child is reluctant to go out. (c) Mother gently nudges child towards door. (d) Child explores outside. (e) Mother and child play happily together.

Star Film:

1. **QUESTION:** Now tell me what you saw in the cartoon
   RESPONSE: triangles
2. **QUESTION:** Tell me what happened
   RESPONSE: they're dancing about

Coaxing Film:

1. **QUESTION:** Now tell me what you saw in the cartoon
   RESPONSE: twisting around
2. **QUESTION:** Tell me what happened
   RESPONSE: no response
3. **QUESTION:** Let's pretend the big triangle and the little triangle are people. If they were people, what were they doing?
   RESPONSE: walking
4. **QUESTION:** what do you think the people were feeling?
   RESPONSE: happy…cause theyíre going to see their friends
5. **QUESTION:** why is the big triangle pushing the little one?
   RESPONSE: to get him out
6. **QUESTION:** What are they doing now that they're outside the square?
   RESPONSE: no
7. **QUESTION:** what kind of person is the big triangle?
   RESPONSE: dad
8. **QUESTION:** What kind of person is the little triangle?
   RESPONSE: a little sister

Surprise Film:

1. **QUESTION:** Now tell me what you saw in the cartoon
   RESPONSE: no response
2. **QUESTION:** Tell me what happened
   RESPONSE: so the little one was knocking on the door and the big one came went down outside
3. **QUESTION:** Let's pretend the big triangle and the little triangle are people. If they were people, what were they doing?
   RESPONSE: playing so they have fun
4. **QUESTION:** what do you think the people were feeling?
   RESPONSE: pink (?)
5. **QUESTION:** Look the little triangle knocks on the square and then goes away-why do you think it does that?
   RESPONSE: so the big one doesn't know where he is
6. **QUESTION:** What are they doing now that they're inside the square?
   RESPONSE: they fight cause the big one and the tiny one is tricking the big one and the other one the big one got cross
7. **QUESTION:** what kind of person is the big triangle?
   RESPONSE: a dad
8. **QUESTION:** What kind of person is the little triangle?
   RESPONSE: a little sister

Fig. 2. Sample transcript of a child's description of the Castelli animations.

Table 1
Examples of proposition types

Proposition type	Example	Reliability
Undirected actions (Act)	He's messing about	0.79
Action between animates (ActA)	The big one's trying to hit the little one	0.86
Action on the rectangle (ActR)	He's closing it up	0.65
Relation with the rectangle (RelR)	He went in the square	0.48
Intentional (Int)	The triangle knew the way	0.74
Neutral (Neu)	There's a small triangle	0.90

Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge, a method our group has reported previously (Knickmeyer et al., 2005; Lutchmaya et al., 2002a,b). Amniotic fluid was extracted with diethyl ether. Recovery experiments have demonstrated 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-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.

There were significant differences between boys' and girls' T levels, $t(36) = 7.3$, $P = 0.00$, $d = 2.2$. 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 fT 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 fT. Only 1 girl scored below 0.2 nmol/l, indicating that there was not a strong floor effect. 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 (1.04 and 0.36 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, $d = 2.2$ 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 and 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 et al., 2003). Amniotic estradiol levels were also assayed by the Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge, using the following method.

Amniotic fluid was extracted with diethyl ether. Recovery experiments have demonstrated 95% 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 microtiter 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. There were no significant differences between estrogen levels in boys and girls, $t(19) = 0.64$, $P = 0.53$, $d = 0.22$.

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 et al., 1993). Amniotic AFP levels were also assayed by the Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge. 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% 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 10,199 U/mL and 1.1 at 12,438 U/mL. There was a trend for girls to have higher AFP levels than boys, but this did not reach statistical significance, $t(27) = 1.94$, $P = 0.06$, $d = 0.64$. There is no sex difference in AFP levels in the entirety of the recruitment group.

Sex of child

Boys were coded as 1 and girls were coded as −1 for all analyses.

Gestational age at amniocentesis (weeks)

Levels of fT 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 fT 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. Neither males nor females showed a significant linear relationship between fT and gestational age, $r(21) = 0.08$, $P = 0.72$ and $r(14) = -0.35$, $P = 0.22$ for males and females respectively. No quadratic relationships were apparent.

Sociodemographic variables

Maternal 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 et al., 1991). 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. There were no sex differences in maternal age, $t(29) = 0.18$, $P = 0.86$, $d = 0.05$, or number of siblings, $t(29) = 0.17$, $P = 0.87$, $d = 0.07$. Table 2 summarizes descriptive data for all predictor variables for both males and females.

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

Variable	Girls $n = 14$			Boys $n = 25$			d
	Mean	SD	Range	Mean	SD	Range	
fT (nmol/l) *	0.37	0.18	0.19–0.80	1.04	−0.40	0.13–1.73	2.2
Estrogen (pmol/l)	1041	485	591–1950	948	315	440–1750	0.22
AFP (MU/l)	11.4	3.03	7.60–19.7	9.45	3.01	3.10–14.5	0.64
Gestational age at amino (weeks)	16.6	0.94	15–18	17.0	1.75	14–21	0.28
Maternal age	35.1	4.22	27–40	35.3	4.49	28–43	0.05
Number of siblings	1.07	1.00	0–3	1.13	1.06	0–3	0.07

* $p < 0.01$.

Results

The first set of analyses provided basic descriptive statistics. Table 3 presents means, standard deviations, and ranges for outcome variables for each sex separately. Kolmogorov–Smirnov tests indicated that ActR was significantly skewed. ActR scores were logged. Where the original score was 0, the logged score was recorded as −2.00. Transformation reduced skewness but increased kurtosis. Kolmogorov–Smirnov tests indicated that ActrR was still significantly skewed. We investigated the distribution of scores to determine whether there were floor effects for any of the variables. ActR showed a strong floor effect with 50% of children not producing the category. The other categories showed small floor effects. For number of mental state terms, 20% of children used no such terms at all. For ActA, 13.5% of children did not produce this category of proposition. For RelR, 22% of children did not produce this category. For Int, 17% of children did not produce this category. ActR was discarded from the subsequent analyses.

Sex differences were tested for using a *t* test, and equal variances were not assumed. Females used significantly more affective state terms than males, $t(19) = -2.17$, $P = 0.04$, $d = 0.82$. Males produced significantly more neutral propositions, $t(33) = 2.01$, $P = 0.05$, $d = 0.63$. There was a trend for females to produce more intentional propositions, $t(21) = -1.90$, $P = 0.07$, $d = 0.62$. Females also used more mental state terms than males, $t(22) = -1.53$, $P = 0.14$, $d = 0.49$. Although the *t* test for this comparison does not meet traditional standards for significance, the effect size is medium, suggesting that sex differences could be apparent in a larger sample. Therefore, these scales were explored further.

Relations between outcome variables and fetal testosterone levels

Hierarchical regression analysis was used to explore the contributions of the predictor variables to variation in the outcome variables. However, it should be kept in mind that, with a sample size of 39, the analysis had limited power. In the first block, any predictor variable that correlated significantly with the outcome variable at $P < 0.2$ was forced 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 4 for correlations between all variables). In the second block, the main effects of fT and child's sex were tested for inclusion with a stepwise analysis. In the third block, 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 outcome variables. Entry criteria was $P < 0.05$; removal criteria was $P > 0.1$.

None of the background predictor variables correlated with the either number of mental state or number of affective state terms used at the $P < 0.2$ level (see Table 4). Therefore, in the first block, the main effects of fT and child's sex were tested for inclusion with a stepwise analysis. In the third block, the interaction of sex and fT was tested for inclusion with a stepwise analysis. For mental state terms, neither sex or fT or the interaction of sex and fT met inclusion criteria for the model. For affective state terms, inclusion of sex in the first stage produced a significant *F* change, *F* change = 6.09, $\beta = -0.38$, $P = 0.02$. This model explains 14% of the variance in frequency of affective state terms. Residual analysis showed acceptable plots and no outliers. Although our results did not suggest that fT played an important role in frequency of mental and affective state terms, because of the correlation between fT and sex, we cannot rule out a role for fT. To further investigate any potential relationship, we analyzed 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. Neither mental state terms nor affective state terms correlated significantly with fT within either sex, $r(25) = 0.003$, $P = 0.99$ and $r(14) = 0.16$, $P = 0.58$ for boys and girls respectively and $r(25) = 0.03$, $P = 0.90$ and $r(14) = -0.18$, $P = 0.53$.

For intentional propositions, the following variables were entered: AFP and gestational age (suppressor). Inclusion of fT in the 2nd stage produced a significant *F* change, *F* change = 7.56, $\beta = -0.44$, $P = 0.01$. 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 implies that the relationship between fT and quality of social relationships is the same for boys and girls. Those with lower fT levels produced more intentional propositions. The only significant predictor in the final model was fT (see Table 5). Residual analysis showed acceptable plots and no outliers. To further investigate whether the previous result might be due to a sex difference (not necessarily involving testosterone), we analyzed the relationship between Int and fT within each sex. It should be kept in mind that this reduces the sample size by half and therefore reduces the power of the analysis. Within females, there was little correlation between fT levels and Int, $r(13) =$

Table 3
Means, standard deviations, and ranges for outcome variables by sex

Variable	Boys $n = 25$			Girls $n = 14$			*d*
	Mean	SD	Range	Mean	SD	Range	
Number of propositions	27.9	9.03	9–41	26	7.82	15–43	0.21
Mental state terms	0.07	0.07	0.00–0.33	0.11	0.09	0.00–0.26	0.50
Affective state terms	0.04	0.05	0.00–0.14	0.09	0.07	0.00–0.22	0.82 *
Act	0.22	0.12	0.00–0.46	0.22	0.11	0.10–0.42	0.00
ActA	0.11	0.10	0.00–0.33	0.12	0.11	0.00–0.43	0.00
ActR	0.03	0.04	0.00–0.15	0.04	0.04	0.00–0.10	0.25
RelR	0.07	0.07	0.00–0.26	0.08	0.07	0.00–0.23	0.14
Int	0.06	0.07	0.00–0.33	0.11	0.09	0.00–0.26	0.62
Neu	0.45	0.15	0.24–0.89	0.37	0.10	0.19–0.53	0.63 *

Note. All variables except number of propositions represent total counts divided by the number of propositions.
* $p < 0.05$.

Table 4
Correlation matrix showing relationships between the independent variables for all subjects of both sexes (n = 33–39)

	Mental state terms	Affective state terms	Int	Neu	fT (nmol/l)	Child's sex	Gestational age	AFP (MU/l)	Estrogen (pmol/l)	Maternal age	Number of siblings
Mental state terms											
Affective state terms	−0.09										
Int	0.70 **	0.35 *									
Neu	−0.11	−0.24	−0.05								
fT (nmol/l)	−0.31	−0.28	−0.43 **	0.23							
Child's sex	−0.26	−0.38 *	−0.32	0.29	0.70 **						
Gestational age	0.09	0.13	−0.13	0.29	0.13	0.16					
AFP (UM/l)	0.15	−0.02	0.32	−0.25	−0.11	−0.30	−0.70 **				
Estrogen (pmol/l)	−0.06	0.11	0.18	0.10	0.03	−0.12	−0.14	0.32 *			
Maternal age	0.08	−0.13	0.07	−0.28	0.06	0.03	−0.51 **	0.22	−0.14		
Number of siblings	0.17	−0.09	0.01	−0.19	0.06	0.03	−0.23	0.19	−0.02	0.41 *	

Note. n varies due to missing data for some participants. Correlations are Pearson correlations.
* $p < 0.05$.
** $p < 0.01$.

−0.14, $P = 0.64$. Within males, fT levels and Int did correlate, $r(24) = -0.37$, $P = 0.07$ (one tailed $P = 0.04$).

For neutral propositions, the following variables were entered: gestational age, AFP, and maternal age. Inclusion of fT in the 2nd stage produced a significant F change, F change = 4.28, $\beta = 0.36$, $P = 0.05$. 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 does 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 higher fT levels produced more neutral propositions. The only significant predictor in the final model was fT (see Table 5). Residual analysis showed acceptable plots and no outliers. To further investigate whether the previous result might be due to a sex difference (not necessarily involving testosterone), we analyzed the relationship between Neu 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. Neu did not correlate significantly with fT within either sex, $r(24) = 0.05$, $P = 0.81$ and $r(13) = -0.04$, $P = 0.90$ for boys and girls respectively.

Table 5
Hierarchical regression analysis, final model: sexes combined

Dependent variable	R^2	Predictors	B	SE B	Sig
Int	0.28	Constant	−0.21	0.34	0.54
		AFP (MU/l)	0.01	0.01	0.35
		Gestational Age	0.01	0.01	0.65
		Estrogen (pmol/l)	0.07	0.08	0.42
		fT (nmol/l)	−0.08	0.03	0.01 **
Neu	0.23	Constant	0.57	0.62	0.36
		AFP (MU/l)	−0.01	0.01	0.60
		Gestational Age	0.00	0.02	0.86
		Maternal Age	−0.01	0.01	0.31
		fT (nmol/l)	0.11	0.05	0.05 *

* $p \le 0.05$.
** $p \le 0.01$.

Discussion

In this study, we examined whether fT was related to the tendency to interpret ambiguous visual stimuli in intentional and human terms. This test was used as a measure of mental state attribution, a component of empathy. Typically, developing children were presented with a series of films featuring shapes whose movements were designed to elicit theory of mind attributions and recorded the children's descriptions. Their narratives were analyzed for the frequency of mental and affective state terms and classified all the propositions in their narratives according to the criteria set out by Bowler and Thommen (2000). We predicted that females would use more mental and affective state terms than males. We also predicted that females would make more intentional propositions, as classified by Bowler and Thommen (2000); intentional propositions include propositions describing actions between animate or agentive entities. We also predicted that variation in fT levels would account for the predicted sex differences. In general, our predictions were supported. Our results implicate fT in human social development. They are also compatible with the 'empathizing–systemizing' theory of sex differences (Baron-Cohen, 2003) since this predicts that females in the general population will score higher than males on a test of empathy. Finally, in isolating fT as related to some aspects of empathy, the study provides further support for the hypothesis that fT is a risk factor for autism.

Mental and affective state terms

Females used mental state terms more frequently than males, but the relationship did not meet traditional standards of significance. The effect size for the relationship was medium ($d = 0.49$), according to Cohen's guidelines, and was similar to the effect sizes seen in our earlier studies. The majority of psychological studies demonstrate moderate effect sizes (i.e. $d = 0.5$) (Eagly, 1995). This suggests that the primary reason we did not observe a significant difference may have been insufficient sample size. Assuming that the mean difference seen in this sample accurately reflects the mean difference in the population, a sample size of approximately 80 would be required to give the model a power of 0.6. Neither child's sex nor fT was included in the final regression model. Once again, this may indicate that the analysis was underpowered. Ideally, regression analyses should include at least 50 individuals. Although we invited approximately 60 families to take part in the study, difficulties in attending the laboratory for testing meant that only 38 families participated.

It should also be kept in mind that mental state terms were relatively infrequent and that 20% of the children produced no such terms at all. In order to determine whether the films were effective at eliciting mental state attributions, we compared the frequency of such terms in our sample to those found in similar studies. Bowler and Thommen (2000) reported a near floor effect on mental state terms for all groups in their study using the original Hieder and Simmel film, suggesting that the film was not a good instrument for eliciting the use of mental state terms. For the category 'think/know,' they reported mean numbers (SDs) per group as 0.18 (0.71), 0.10 (0.32), 0.00, and 0.18 (0.60) for those with autism, chronological age (CA)-matched, verbal mental age (VMA)-matched, and IQ-matched controls, respectively. For the category 'want,' they reported mean numbers (SDs) per group as 0.36 (0.84), 0.00, 0.36 (0.63), and 0.09 (0.30) for those with autism, CA-matched, VMA-matched, and IQ matches respectively. In contrast, in our study, the mean mental state terms (includes both think/know and want) for the entire group were 2.0 (1.7). Abell et al. (2000) reported mean number of mentalizing descriptions as 1.73 (1.03) for typically developing 8-year-olds and 3.57 (0.65) for typical adults (note that the films we used are a subset of those used by Abell et al.). Klin (2000) reported 13.6 (10.7) for the cognitive index on their film.

Comparison is difficult because each study used different methods and produced narratives of varying length. If we divide the reported mean for each study by the number of propositions (or in the case of Abell et al. (2000) the number of explanations), we find that, at 0.08, the children in our study produced terms more frequently than the best scoring group in Bowler and Thommen (2000) (0.02) but far fewer than the typically developing groups in Abell et al. (2000) and Klin (2000), 0.32, 0.90, and 0.30 respectively. Abell et al. (2000) found that use of mentalizing language increased from age 8 to adulthood. The children in our sample were age 4. It is therefore possible that, if the children were retested at a later age, they would use a greater number of mental state terms. Sex differences and relations to fT might be more apparent at that age. We will have the opportunity to retest this sample when they are older.

Our results thus suggest that, with a larger sample or at a later age, a sex difference may exist. However, we cannot speculate on whether this difference would be related to fT. Sex differences may be produced by many factors, both biological and social. Within-sex correlations of fT and mental state terms were not significant. It is true that the sample size is small. However, significant results have been obtained in amniocentesis studies with similar sample sizes. Grimshaw et al. (1995) found significant within-sex correlations for amniotic fT and mental rotation speed for both girls and boys, $r(12) = 0.67$ and $r(13) = -0.62$. r values were much smaller in the current study, suggesting that even in a larger sample no strong relationship between fT and mental state terms would be observed. However, sex differences in mental rotation are larger ($d = 0.9$) than those we measured for mental state terms. If the correlation of 0.16 in girls and 0.18 in boys seen in this study reflected an actual population correlation, sample sizes of 110 and 150, for girls and boys respectively, would be needed for the model to have a power of 0.6. Even if within-sex variations in fT do not contribute to differences in the frequency of mental state terms, this does not rule out a role for fT in producing a sex difference on this variable. fT may only exert its effects on mental state terms at very high doses.

For use of affective state terms, a significant sex difference was seen, with girls using such terms more frequently than males. Sex was the only significant predictor in the final models of our regression analyses for both male and female items; fT was excluded from the analysis. Again, this may indicate that the analysis was underpowered. It is also possible that sex differences in the use of affective state terms are the result of other variables, either biological or social. Dunn et al. (1991) reported that children who grew up in families in which they engaged in conversations about feelings were better able to explain the feelings and actions of puppet characters when tested 7 months later. Their results are compatible with a role for such conversation in promoting conceptual development, although because the child is an active participant in such conversations, the relationship could also reflect a common underlying ability. Dunn et al. (1987) reported that mothers talk more about feelings to their daughters, although Fivush (1989) notes that this may depend on the type of emotion. Maternal age and number of siblings were not significantly related to use of affective state terms in the current study, but these variables are quite broad. When looking within sex, fT was not significantly correlated with scores on either scale, and r values were low. Our results suggest that amniotic fT, measured at this period, does not account for individual variation in use of affective state terms, but, because of the sample size issues, we cannot rule out a role for fT in the differences seen between boys and girls.

Our results for mental and affective state terms support the 'empathizing–systemizing' (E–S) theory of sex differences (Baron-Cohen, 2003) in that boys performed at a lower level to girls. Our results are also consistent with the 'extreme male brain' theory of autism (Baron-Cohen, 2003) in that, from other studies, we know that individuals on the autistic spectrum perform even lower than typical males on the moving geometric shapes task. However, we must acknowledge that the observed sex differences could have arisen from differences in socialization. Klin (2000) found that individuals with autism used fewer mental and affective state terms when describing films similar to the ones we used, and Baron-Cohen et al. (1986) found that children with autism used fewer mental state terms when performing a picture-sequencing task. It should be noted that Bowler and Thommen (2000) did not find a difference between individuals with autism and controls when describing the Heider and Simmel (1944) film, but this probably reflects the very low frequency of such terms in all groups. Abell et al. (2000) found that children with autism did not provide fewer mental state explanations than controls with moderate learning disability, but they did find that the children with autism were less accurate. We were unable to meaningfully test whether normally developing 4-year-olds showed sex differences in the accuracy of their descriptions as overall accuracy was very low. Abell et al. (2000) classified accuracy for the ToM sequences according to the following criteria.

Surprising:

2—any mention of the little triangle tricking, surprising the big triangle; hiding; or playing hide and seek
1—description which gives part of the story but misses the critical point (see above)
0—description which only gives a minor part of action, i.e. knocking on the door, or does not relate to any of the events in the sequence.

Coaxing:

2—description that conveys the child's reluctance to go out and the mother's attempts to get the child out, e.g., persuading
1—partially correct description focusing on one aspect of the story or one character only, e.g., child does not want to go out or mother is pushing child to go out
0—actions that do not relate to the events or relate to a minor aspect of the sequence only, e.g., dancing together, or unrelated description.

Only 4 children in our sample achieved a score of 2 on the surprising sequence. No child achieved a score of 2 on the coaxing sequence.

Intentional propositions

Bowler and Thommen (2000) found that children with autism were significantly less likely to use propositions describing actions between animates (ActA) than a variety of control groups. They interpreted this as reflecting a difficulty in perceiving the goal-directedness of an action when it is embedded in a complex time-pressured context. They suggest that, for the child with autism, someone engaging in an action is doing just that, and the consequences and directedness of the action are less likely to be taken into consideration when describing an event. In keeping with the extreme male brain theory, we predicted that boys would produce fewer ActA propositions than girls. There was no significant sex difference on this measure, and the effect size was very small, indicating that even in a larger sample no differences would be apparent.

However, we did find a significant sex difference in the frequency of intentional propositions (Int), with girls using more than boys. These are intentional propositions not involving action with animates or action with the rectangle. In practice, these included many propositions describing mental and affective states but also included a broader range of utterances. There was a main effect of fT on this variable 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 fewer intentional propositions. 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 24 boys in the sample, but only 13 girls). A sample size of approximately 26 would be required to give the model a power of 0.6, 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.

The results of this laboratory study are consistent with the inverse relationship we observed between fT and quality of social relationships assessed by maternal questionnaire (Knickmeyer et al., 2005). Causal interpretations cannot, of course, be drawn from this correlational study, but the observed correlations are consistent with the hypothesis that fT reduces social attributions in both boys and girls, presumably as a function of altered neural development. Another explanation would be that relations between fT and social cognition are mediated by other variables not measured in this study or that fT is serving as an index for an unknown third variable. We also observed a significant sex difference for frequency of neutral propositions, with boys producing more of these than girls. Bowler and Thommen (2000) reported that children with autism produce more neutral propositions, particularly in comparison with IQ matched controls, so this result is in keeping with the extreme male brain theory. They interpreted this as a pragmatic impairment. Although a few such statements are necessary to create a coherent story, children with autism peppered their accounts with these statements which describe non-action-related aspects of the film. fT was directly related to the number of neutral propositions when the group was examined as a whole.

Limitations

All amniocentesis studies are limited by the problems inherent in studying fetal endocrinology. We have assumed that amniotic T levels accurately represent serum levels and brain exposure, but 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 determine whether and how potent T's effects may be.

Secondly, 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 not always available for all children. A final limitation of research using this method is that a truly random sample cannot be collected since one can only include individuals whose mothers 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 fT levels are different in mothers who undergo amniocentesis compared to those that do not because, within this group, no relationship was seen between fT and maternal age, alpha-fetoprotein level, paternal age, or parental education level (Lutchmaya, 2000).

Scarr and McCartney (1983) and others have also pointed out that there is no such thing as a purely biological or purely environmental contribution to behavior. Addressing all potential factors and their interactions would require an immense sampling size. Our focused relatively small-scale studies are, we argue, nevertheless valuable in demonstrating that fetal testosterone potentially has long-term effects on social relationships and social cognition.

Acknowledgments

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. None of the authors have a significant financial arrangement, affiliation with, or potential bias against any product or services used or discussed in this paper. 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. SBC was also supported by the Medical Research Council (MRC) UK. We thank Fulvia Castelli for providing the visual stimuli. We are grateful to Ian Goodyer, Sally Wheelwright, Svetlana Lutchmaya, Rosa Hoekstra, and Chloe Miller for their support and discussion and to Emma Chapman and Bonnie Ayeung for assistance with the larger study of which this is one part. Finally, we are indebted to the families who participated in this study and who have repeatedly participated as is essential in prospective longitudinal research.

References

Abell, F., Happe, F., Frith, U., 2000. Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. Cogn. Dev. 15, 1–16.

Altman, D.G., 1991. Practical Statistics for Medical Research. Chapman and Hall, London.

APA, 1994. DSM-IV Diagnostic and Statistical Manual of Mental Disorders, 4th ed. American Psychiatric Association, Washington, DC.

Arnold, A.P., 1996. Genetically triggered sexual differentiation of brain and behavior. Horm. Behav. 30, 495–505.

Arnold, A.P., Gorski, R.A., 1984. Gonadal steroid induction of structural sex differences in the central nervous system. Annu. Rev. Neurosci. 7, 413–442.

Baron-Cohen, S., 2002. The extreme male brain theory of autism. Trends Cogn. Sci. 6, 248–254.

Baron-Cohen, S., 2003. The Essential Difference: The Male and Female Brain, and The Riddle of Autism. Penguin, London.

Baron-Cohen, S., Hammer, J., 1997. Is autism an extreme form of the male brain? Advances Infancy Res. 11, 193–217.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1986. Mechanical, behavioural and Intentional understanding of picture stories in autistic children. Br. J. Dev. Psychol. 4, 113–125.

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. J. Child Psychol. Psychiatry 38, 813–822.

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

Baron-Cohen, S., Lutchmaya, S., Knickmeyer, R.C., 2004. Prenatal Testosterone in Mind: Amniotic Fluid Studies. MIT Press, Cambridge, MA.

Bartsch, K., Wellman, H.M., 1995. Children Talk About the Mind. Oxford Univ. Press, Oxford.

Bishop, D.V.M., 1990. Handedness and Developmental Disorder. Blackwell, Oxford.

Bishop, D.V.M., 1998. Development of the children's communication checklist (CCC): a method for assessing qualitative aspects of communicative impairment in children. J. Child Psychol. Psychiatry 6, 879–891.

Bowler, D.M., Thommen, E., 2000. Attribution of mechanical and social causality to animated displays by children with autism. Autism 4, 147–171.

Breedlove, S.M., 1994. Sexual differentiation of the human nervous system. Annu. Rev. Psychol. 45, 389–418.

Castelli, F., Happe, F., Frith, U., Frith, C., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 12, 314–325.

Collaer, M.L., Hines, M., 1995. Human behavioural sex differences: a role for gonadal hormones during early development? Psychol. Bull. 118, 55–107.

Crick, N.R., Casas, J.F., Mosher, M., 1997. Relational and overt aggression in pre-school. Dev. Psychol. 33, 579–588.

De Vries, G.J., Simerly, R.B., 2002. Anatomy, development, and function of sexually dimorphic neural circuits in the mammalian brain. In: Pfaff, D.W., Arnold, A.P., Etgen, A., Fahrbach, S., Rubin, R.T. (Eds.), Hormones, Brain and Behavior, vol. 4. Academic Press, New York, pp. 137–191.

De Vries, G.J., Rissman, E.F., Simerly, R.B., Yang, L., Scordalakes, E.M., Auger, C.J., Swain, A., Lovell-Badge, R., Burgoyne, P.S., Arnold, A.P., 2002. A model system for study of sex chromosome effects of sexually dimorphic neural and behavioral traits. J. Neurosci. 22, 9005–9014.

Dunn, J., Bretherton, I., Munn, P., 1987. Conversations about feeling states between mothers and their young children. Dev. Psychol. 23, 132–139.

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 Dev. 62, 1352–1366.

Eagly, A.H., 1995. The science and politics of comparing women and men. Am. Psychol. 50 (3), 145–158.

Finegan, J.K., Bartleman, B., Wong, P.Y., 1989. A window for the study of prenatal sex hormone influences on postnatal development. J. Genet. Psychol. 150, 101–112.

Fitch, R.H., Bimonte, H.A., 2002. Hormones, brain, and behavior: putatative biological contributions to cognitive sex differences. In: McGillicuddy-DeLisi, A.V., DeLisi, R. (Eds.), Biology, Society, and Behavior: The Development of Sex Differences in Cognition. Ablex, Greenwich.

Fitch, R.H., Denenberg, V.H., 1998. A role for ovarian hormones in sexual differentiation of the brain. Behav. Brain Sci. 21, 311–352.

Fivush, R., 1989. Exploring sex differences in emotional content of mother–child conversations about the past. Sex Roles 20, 675–691.

Geary, D.C., 1998. Male, Female: The Evolution of Human Sex Differences. American Psychological Association, Washington, DC.

Geary, D.C., 2002. Sexual selection and sex differences in social cognition. In: McGillicuddy-DeLisi, A.V., DeLisi, R. (Eds.), Biology, Society, and Behavior: The Development of Sex Differences in Cognition. Ablex, Greenwich.

Goy, R.W., McEwen, B.S., 1980. Sexual Differentiation of the Brain. The MIT Press, Cambridge, MA.

Grimshaw, G.M., Sitarenios, G., Finegan, J., 1995. Mental rotation at 7 years: relations with prenatal testosterone levels and spatial play experiences. Brain Cogn. 29, 85–100.

Halpern, D.E., 1992. Sex Differences in Cognitive Abilities. Erlbaum, Hillsdale, NJ.

Heider, F., Simmel, M., 1944. An experimental study of apparent behaviour. Am. J. Psychol. 57, 243–259.

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

Hines, M., Ahmed, S.F., Hughes, I.A., 2003. Psychological outcomes and gender-related development in complete androgen insensitivity syndrome. Arch. Sex Behav. 32, 93–101.

ICD-10, 1994. International Classification of Diseases, 10th ed. World Health Organisation, Geneva, Switzerland.

Joliffe, T., Baron-Cohen, S., 1997. Are people with autism of Asperger's Syndrome faster than normal on the Embedded Figures Task? J. Child Psychol. Psychiatry 38, 527–534.

Kimura, D., 1999. Sex and Cognition. The MIT Press, Cambridge, MA.

Klin, A., 2000. Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: the Social Attribution Task. J. Child Psychol. Psychiatry 7, 831–846.

Klopper, A., 1970. Steroids in amniotic fluid. Ann. Clin. Res. 2, 289–299.

Knickmeyer, R.C., Baron-Cohen, S., Raggatt, P., Taylor, K., 2005. Foetal testosterone, social relationships, and restricted interests in children. J. Child Psychol. Psychiatry 46, 198–210.

Lutchmaya, S., 2000. Foetal testosterone and social-communicative development. Ph.D. Dissertation, University of Cambridge.

Lutchmaya, S., Baron-Cohen, S., Raggatt, P., 2002a. Foetal testosterone and eye contact in 12 month old infants. Infant Behav. Dev. 25, 327–335.

Lutchmaya, S., Baron-Cohen, S., Raggatt, P., 2002b. Foetal testosterone and vocabulary size in 18- and 24-month-old infants. Infant Behav. Dev. 24, 418–424.

MacLusky, N., Naftolin, F., 1981. Sexual differentiation of the central nervous system. Science 211, 1294–1303.

Montogomery, D.E., Montogomery, D.A., 1999. The influence of movement and outcome on young children's attributions of intention. Br. J. Dev. Psychol. 17, 245–261.

Nagami, M., McDonough, P.G., Ellegood, J.O., Mahesh, V.B., 1979. Maternal and amniotic fluid steroids throughout human pregnancy. Am. J. Obstet. Gynecol. 134, 674–680.

Robinson, J.D., Judd, H.L., Young, P.E., Jones, D.W., Yen, S.S.C., 1977. Amniotic fluid androgens and estrogens in midgestation. J. Clin. Endocrinol. 45, 755–761.

Rutter, M., 1978. Language disorder and infantile autism. In: Rutter, M., Schopler, E. (Eds.), Autism: A Reappraisal of Concepts and Treatment. Plenum, New York.

Scarr, S., McCartney, K., 1983. How people make their own environments: a theory of genotype→environment effects. Child Dev. 54, 424–435.

Scott, F., Baron-Cohen, S., Bolton, P., Brayne, C., 2002. Prevalence of autism spectrum conditions in children aged 5–11 years in Cambridgeshire, UK. Autism 6, 231–237.

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

Springer, K., Meier, J.A., Berry, D., 1996. Nonverbal bases of social perception: developmental change in sensitivity to patterns of motion that reveal interpersonal events. J. Nonverbal Behav. 20, 199–211.

Wathen, N.C., Campbell, D.J., Kitau, M.J., Chard, T., 1993. Alphafetoprotein levels in amniotic fluid from 8 to 18 weeks of pregnancy. Br. J. Obstet. Gynaecol. 100, 380–382.

Williams, C.L., Meck, W.H., 1991. The organizational effects of gonadal steroids on sexually dimorphic spatial ability. Psychoneuroendocrinology 16, 155–176.

Wing, L., 1981. Asperger Syndrome: a clinical account. Psychol. Med. 11, 115–130.