Prenatal testosterone and sexually differentiated childhood play preferences: a meta-analysis of amniotic fluid studies

Gareth Richards[1] · Wendy V. Browne[2]

Abstract

Sexually differentiated behaviour appears to emerge from a complex interaction of biological and socio-cultural factors, with prenatal exposure to steroid hormones such as testosterone thought to play a key role. Due to large sex differences being present from a very early age, much research has focussed on the influence these hormones may have on play preferences during childhood. We present an overview of the literature and a random-effects meta-analysis linking amniotic testosterone with sexually differentiated play preferences (k = 9, n = 493). The overall effect size estimate was in the theory-consistent direction (i.e., with higher levels of testosterone associated with more male-typical play preferences), though not statistically significant ($r = 0.082$, $p = 0.274$). However, after three hypothesised missing studies were imputed via the trim and fill procedure, a significant correlation emerged ($r = 0.166$, $p = 0.014$). Nevertheless, one sample was observed to exert a particularly large influence on the outcome of the analysis. Notably this was the second biggest sample and related to the largest effect size estimate. Though far from conclusive, the overall findings are consistent with the idea that individual differences in prenatal testosterone within the typical range predict sexually differentiated play preferences in early life. However, these effects may be small in magnitude and appear to vary considerably across studies.

Keywords Amniocentesis · Meta-analysis · Play · Sex differences · Testosterone

Introduction

Sex differences are present for a range of behaviours that emerge during early human development (Hines, 2004). One of the most widely reported and consistently replicated findings is that male and female children differ, on average, in their play preferences. Although we do not have a full understanding of the exact origins and mechanisms by which these preferences develop, it appears likely that a complex interaction between biological and environmental influences occurs. Interestingly, sex differences reminiscent of those reported in humans have also been noted in other mammalian species, an observation that implies a common evolutionary origin (LaFreniere, 2011).

✉ Gareth Richards
gareth.richards@ncl.ac.uk

1 School of Psychology, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

2 Faculty of Education, University of Cambridge, Cambridge, UK

Sex differences in human play behaviour manifest very early in life as a male inclination towards 'rough-and-tumble' (e.g., play fighting), and preferences by both boys and girls for same-sex play partners and stereotypically sex-typical toys (e.g., vehicles for boys, dolls for girls). The magnitude of the sex difference for play behaviours appears to be particularly large for toy preferences, which is also larger than the size of the sex differences found in most areas of cognition and personality (Hines, 2010). A recent meta-analysis (k = 75) by Davis and Hines (2020) confirmed that boys prefer male-typical toys ($d = 1.83$) and girls prefer female-typical toys ($d = 1.60$), that both boys ($d = 3.48$) and girls ($d = 1.21$) prefer gender-typical toys over gender-atypical toys, and that these effects are very large in magnitude. It was also noted that boys' preference for male-typical toys and girls' preference for female-typical toys increases with age. Todd et al. (2018) reported similar findings for a meta-analysis (k = 16) of children's toy choices during free play activities and noted that the effects were not moderated by a variety of social/cultural factors including the presence of an adult, the study context, the gender inequality index of the country in which the study was conducted, and the inclusion

or exclusion of gender-neutral toys. It was also noted that boys play more with male-typical toys in lab studies than when observed at home, with the authors speculating that the testing context may make sexually differentiated behaviours more salient, particularly for boys, thereby influencing toy preferences. Furthermore, compared with findings from older studies, girls in more recent studies played less with both male-typical and female-typical toys whereas boys in more recent studies played less with male-typical (but not female-typical) toys. This finding, however, contrasts somewhat with the larger meta-analysis by Davis and Hines (2020), which did not observe a significant association between the year of study and the magnitude of the observed sex difference. Taken together though, the pattern of findings reported from this literature is generally consistent with both biological and environmental influences contributing to the development of childhood play preferences.

As previously indicated, children also show reliable sex differences in the degree to which they engage in active play, with boys being more inclined towards rough-and-tumble and aggressive/contact play compared to girls (Hines, 2013a). Notably, this finding has been replicated across cultures (DiPietro, 1981; Maccoby, 1998; Smith & Connolly, 1980), as well as in other primate species (Fagen, 1981; Smith, 1982). A further sex difference has been found in playmate preferences, with boys and girls both preferring same-sex playmates more than other-sex playmates (Hines & Kaufman, 1994; Maccoby, 1988; Maccoby & Jacklin, 1987; Martin & Fabes, 2001). This difference has been found to be particularly strong when children are engaged in unstructured activities and when adults are not present (Thorne, 2001). Despite individual variation among each type of play, sex differences have been found consistently, thereby warranting attention to the invariably complex interaction between specific social and biological factors involved in their development.

Testosterone and sexual differentiation

One biological factor that has been of considerable research interest is prenatal testosterone, with male foetuses being exposed to markedly higher levels compared to females. In male foetuses, the main source of androgens is the testes, with a smaller amount also derived from the adrenal glands; in females the main source is the adrenal glands, with a negligible amount additionally being produced by the ovaries (Martin, 1985). There is a steep rise in testosterone in males beginning around gestational week 7 when the SRY gene first initiates development of the foetal testes (Nassar & Leslie, 2021). Maximal sexual differentiation then occurs toward the early-mid second trimester. Notably, Abramovich (1974) reported that testosterone concentrations in umbilical cord plasma are 9 times higher in male than female foetuses

measured between gestational weeks 12 and 18, and that this difference is much larger than that observed at term. As experimental animal research has revealed a linear effect of androgens in which greater exposure results in more male-typical behaviours (Hines, 2011), it is plausible that androgen exposure during critical periods of brain development in humans can contribute not only to between-sex differences but also to within-sex differences. However, as it is clearly unethical to manipulate hormone levels in human pregnancies for research purposes, there is a necessary reliance on indirect methods (for a review, see Cohen-Bendahan et al., 2005). These approaches primarily consist of studies that relate normal variability in prenatal testosterone to later sexually differentiated behaviours, as well as studies that examine individuals who experience atypical hormonal environments due to endocrine disorders.

Play preferences and atypical androgen exposure

Rodent models show that the development of a male phenotype requires not only testosterone itself, but also functioning androgen receptors upon which to act (Sato et al., 2004). In humans, this mechanism can be investigated by studying individuals with complete androgen insensitivity syndrome (CAIS). Individuals with CAIS have a male karyotype (i.e., 46XY) but do not possess functioning androgen receptors. Consequently, despite normal (or even elevated) production of testosterone, the hormone cannot affect developing tissues, and so the condition results in female-typical development both in terms of physical phenotype and behavioural outcomes. It is therefore relevant to note that sexually differentiated play preferences are usually female-typical in 46XY females with CAIS (Hines et al., 2003). However, in the current context, studies of CAIS inform us only of the developmental trajectories that can occur in the (effective) absence of testosterone. Furthermore, individuals with CAIS are assigned female and reared as such. Therefore, the effects of androgen insensitivity and those of socialisation as a female cannot be easily untangled (Jordan-Young, 2010). To better understand the effects of testosterone itself, we therefore turn to studies of congenital adrenal hyperplasia (CAH).

In approximately 90% of cases of CAH, there is a deficiency of the 21-hydroxylase (21-OH) enzyme. This results in low concentrations of cortisol and high concentrations of adrenal androgens beginning at approximately the seventh week of gestation (Merke & Bornstein, 2005). Girls with CAH show elevated male-typical childhood play behaviour (Berenbaum & Hines, 1992; Hines et al., 2004; Meyer-Bahlburg et al., 2004), with an increase in the amount of time spent playing with boys' toys and a decrease in the amount of time spent playing with girls' toys when compared with unaffected female relatives (Berenbaum & Hines, 1992).

Furthermore, a 'dose–response' effect has been documented, with more interest in male-typical toys being found in girls with more severe forms of CAH (Nordenström et al., 2002; Servin et al., 2003). This observation corroborates animal research documenting a linear effect of androgens in which greater exposure results in more male-typical behaviours (Hines, 2011). However, findings have not always been consistent. For instance, Berenbaum and Snyder (1995) found that, compared to unaffected female relatives, girls with CAH did not prefer boys as playmates, but did prefer male-typical toys and activities, whereas Hines and Kaufman (1994) found that, compared to unaffected relatives, girls with CAH did not differ in their amount of rough-and-tumble play, but did show a preference for male playmates. The null findings for rough-and-tumble could result from the play context (Hines, 2004); when playing with girls with CAH, boys may not have been responsive due to a general avoidance of rough-and-tumble play with girls. This is consistent with the study by Pasterski et al. (2011), which found that girls with CAH chose playmates engaged in a masculine activity, regardless of the sex of the playmate (i.e., their choices were driven more by the toys than by the sex of the playmate).

Although male-typical play preferences in girls with CAH are consistent with the idea that androgen excess during the prenatal period promotes bipotential areas of the brain to develop in a male-typical direction (Pasterski, 2008), such findings could also be explainable by socialisation. Nevertheless, Pasterski et al. (2005) reported that masculinised toy preferences in girls with CAH were present despite both mothers and fathers providing more encouragement of female-typical play to their CAH-affected daughters than to their CAH-unaffected daughters. However, a later study by Wong et al. (2013) reported that parents encouraged *more* male-typical and *less* female-typical toy play in their daughters with CAH compared to their unaffected daughters, and that these encouragement patterns partially mediated the association between CAH status and sex-atypical toy play. Although it might be inferred from the findings of this study that socialisation is primarily responsible for sex-atypical play preferences in girls with CAH, it is perhaps more plausible that the parents simply encouraged the play styles and toy choices already favoured by their children (Hines, 2015). Regardless, as the association between CAH and toy preference was only partially rather than fully mediated by parental encouragement, it seems clear that neither prenatal testosterone nor parental encouragement on its own can fully account for the observed effect. Taken together, the evidence from CAH studies suggests that both prenatal testosterone and socialisation play a role in the development of sexually differentiated play preferences. However, it should also be noted that questions have been raised concerning publication bias in this literature (Collaer & Hines, 2020; Hampson, 2016; Richards et al., 2020), and that small sample effects may lead to overestimates of the true population effect size (Open Science Collaboration, 2015).

Prenatal testosterone and play preferences in typical development

Although studying individuals who experience atypical hormonal environments is informative in terms of hormone-behaviour associations, this approach does not result in a perfect experiment (Blakemore et al., 2009). For example, males and females with CAH experience various hormonal abnormalities, including low prenatal concentrations of glucocorticoids, which could influence the developing brain and subsequent behaviour (Nass & Baker, 1991). Likewise, as previously mentioned, most individuals with CAIS are assigned female and reared as such, and so the effects of androgen insensitivity and those of socialisation cannot be untangled easily. Furthermore, any observed behavioural effects may not be directly related to hormonal abnormalities. Rather, there may be other factors which exist alongside the experience of having a long-term illness that could potentially play a role. Ultimately, before firm conclusions can be drawn, findings from studies of individuals with clinical conditions should be considered alongside research that implements alternative paradigms. We therefore turn next to studies that have assayed hormones from maternal serum and amniotic fluid in the idea of determining their influence on subsequently measured sexually differentiated play preferences.

A study by Udry et al. (1995) measured concentrations of total testosterone and sex hormone binding globulin (SHBG) sampled from maternal blood during each of the three trimesters and related these to self-reported questionnaires of gender-related behaviours in 250 female offspring when they were 27–30 years of age. Higher levels of SHBG (but not testosterone) measured during the second trimester (14–26 weeks' gestation) predicted less male-typical behaviour. Hines et al. (2002) later examined sexually differentiated behaviour in 3.5-year-old children and related these behaviours to testosterone measured from maternal blood between gestational weeks 5 and 36 (mean = 16 weeks). A linear relationship between maternal testosterone and sexually differentiated behaviours, including play with sex-typical toys, was found in girls but not boys. Although this finding might indicate that early testosterone exposure plays a role in the development of sexually differentiated play behaviour (at least in girls), it should also be noted that testosterone in the maternal blood did not differ in pregnancies with male versus female foetuses. The findings may therefore be explainable by more masculine mothers socialising their daughters in a male-typical direction (Cohen-Bendahan et al., 2005). Alternatively, these results

could reflect testosterone exposure from different sources in male versus female foetuses. Prenatally, males are predominantly exposed to androgens produced by the testes, with a relatively small amount coming from the adrenal glands; females, however, secrete androgens mainly from the adrenal glands, with a small amount also coming from the ovaries, and their level of production is thought to show a genetic resemblance to that of their mother (Harris et al., 1998). Therefore, masculinised play behaviours in girls that appear to result from higher testosterone production in utero could ultimately reflect a genetic predisposition. Following on from this, the maternal and foetal hormonal environments may not directly influence one another, and a similarity in hormone concentrations could be explainable entirely by shared genetic factors.

Although the findings of Udry et al. (1995) and Hines et al. (2002) differ, both studies indicate that the hormonal environment present during the second trimester appears to be most influential in the development of sexually differentiated behaviour. This is noteworthy because, as mentioned previously, prenatal testosterone levels appear to exhibit maximal differentiation between males and females at this stage. However, as indicated above, questions remain regarding how to interpret hormone-behaviour associations that involve testosterone assayed from the maternal circulation. As a more direct approach to measuring second trimester testosterone would appear to be beneficial, it is fortuitous that clinical amniocenteses are typically performed at this time. Furthermore, a meta-analysis (Baron-Cohen et al., 2015; see supplementary materials of that paper) reported that the testosterone levels present in amniotic fluid extracted via this procedure are much higher for males than females ($d = 1.71$). Amniotic fluid testosterone has also been reported to exhibit maximal sexual differentiation between gestational weeks 12 and 20 (Nagamani et al., 1979; Warne et al., 1977). It is therefore considered likely that testosterone measured in this way is representative of a stage of gestation during which the brain and sexually dimorphic behaviours are malleable to its organisational effects (Knickmeyer et al., 2005a; van de Beek et al., 2004).

Amniotic testosterone and sexually differentiated play have been examined in five independent cohorts. First, Grimshaw et al. (1995) observed no correlation between testosterone and spatial play measured at 7 years of age. Knickmeyer et al. (2005b) later detected no effect for maternally reported masculine/feminine play in 4–5-year-old offspring. Although significant correlations emerged for males and females in a larger study of this cohort (Auyeung et al., 2009) that used the Preschool Activities Inventory (PSAI) at age 8, other researchers in the UK (Spencer et al., 2021; see also Constantinescu, 2009, p. 35) and Germany (Körner, 2018, p. 35) did not replicate the finding. Inconsistent results have also been reported from a Dutch cohort: van de Beek et al.

(2009) found no effect for children's behaviour in structured toy play sessions at 13-months of age, though further longitudinal analyses (Beking, 2018) revealed a different picture. There was no effect in boys but a significant testosterone-by-age interaction in girls was found; more specifically, no associations were observed at 13-months or 2.5-years, but, contrary to established theory, higher testosterone concentrations predicted more female-typical play preferences at 6.5-years.

Aims and hypotheses

The current study presents a meta-analysis of the association between amniotic fluid testosterone and subsequently measured sexually differentiated play preferences. We chose to focus on the specific methodology of amniotic fluid analysis for the following reasons: (1) large sex differences in testosterone assayed from amniotic fluid are reliably detected, whereas this is not the case for other measures such as gestational maternal serum or perinatal umbilical cord blood, (2) amniocenteses are typically performed during the time of gestation at which testosterone exhibits its largest level of sexual differentiation, and (3) participant samples from amniotic fluid studies are more (albeit still not entirely) representative of the general population than those from studies of clinical conditions such as CAH and CAIS. We also focussed on this specific methodology because there has been recent interest in the association between play preferences and testosterone measured using this technique (Spencer et al., 2021) and because amniotic fluid studies have been posited as the most effective method currently available for examining associations between prenatal sex hormone exposure and subsequent behavioural outcomes (Baron-Cohen et al., 2004; van de Beek et al., 2004). Taking account of findings from the wider literature, we predicted that amniotic testosterone would be positively correlated with male-typical play preferences and negatively correlated with female-typical play preferences.

Material and methods

Spencer et al. (2021, p. 7) recently remarked that "both negative and positive results contribute to our understanding of the size, as well as the reliability, of relations between testosterone and gender-related behaviors" and that their results "are of interest not only on their own, but also in the context of prior findings (e.g., for meta-analytic studies)". With this in mind, we searched PubMed (71 hits), Google Scholar (227 hits), ProQuest (46 hits) and Scopus (78 hits) on March 1st 2021 using the terms "amniotic testosterone" AND "play*". Four relevant peer-reviewed journal articles (Auyeung et al., 2009; Knickmeyer et al., 2005b; Spencer

et al., 2021; van de Beek et al., 2009) and two PhD theses (Beking, 2018; Körner, 2018) were identified; to these we added a further peer-reviewed paper (Grimshaw et al., 1995) and an MPhil thesis (Constantinescu, 2009) (see Table 1).

Our first inclusion criterion for the meta-analysis was that studies reported primary data relating to testosterone assayed from amniotic fluid as well as sexually differentiated play preferences measured during infancy/childhood. The age-range of participants included is therefore comparable to that of the recent meta-analysis of sex differences in play preferences reported by Davis and Hines (2020) (i.e., 11 years or younger). We also specified that one or more effect size estimate for the level of association between amniotic testosterone and sexually differentiated play preferences should be available; if effect sizes were not included in the original articles (and could not be calculated from data presented therein), we contacted the author(s) to request this information. To reduce the potential influence of publication bias, we included studies that met the above inclusion criteria regardless of whether they had been published in peer-reviewed journals.

We conducted a random-effects meta-analysis using the R package metafor (Viechtbauer, 2010) to determine the strength and direction of correlation between amniotic testosterone and sexually differentiated play behaviour. We chose to use a random-effects rather than fixed-effects model because the former can account for heterogeneity in effect size estimates beyond that expected by chance/sampling error whereas the latter cannot. In this way, a random-effects model does not assume that there is a single 'true' population effect size for the phenomenon under observation (Borenstein et al., 2009). Effect sizes were standardised across studies as Pearson's r to allow for their direct comparison. However, as Pearson's r is not normally distributed, we converted these to z prior to meta-analysis and then transformed the results back to Pearson's r for ease of interpretation (Borenstein et al., 2009). Point estimates were stratified by sex, and, in cases where the same cohort was examined more than once, the larger/largest available sample was selected. When more than one outcome was considered, the mean of all relevant effect sizes was included. As Grimshaw et al. (1995) reported non-significant correlations but the data are no longer accessible (G. Grimshaw, personal communication), effect sizes for this study were conservatively approximated as $r = 0.000$ (see Hönekopp & Watson, 2011).

There is a risk of bias associated when considering more than one effect size estimate from the same study/article. In the current context, this issue is particularly relevant because all but one of the studies presented separate effect size estimates for male and female subsamples (note that Körner [2018] did not examine females because most females included in the study measured below the detection

threshold for testosterone). To account for this, we conducted both two- and three-level meta-analyses (Konstantopoulos, 2011). The more commonly used two-level meta-analysis controls only for the random effect of the sample from which the estimate is drawn, whereas the three-level meta-analysis nests the random effect of sample within the study/article from which it is drawn. By doing so, the effect size estimate is adjusted to account for any correlation between samples derived from the same source.

Results

The three-level analysis returned an effect size estimate that was noticeably smaller than that of the two-level analysis (Table 2). However, the Akaike Information Criterion (AIC) for the two-level model (AIC = 1.802) was slightly lower than that of the three-level model (AIC = 2.256), suggesting the former to be a better fit for the data. As the likelihood ratio test comparing the two models was not statistically significant, $\chi^2 = 1.546$, $p = 0.214$, we retained the two-level model in subsequent analyses in the interests of parsimony.

The two-level meta-analysis (k = 9, n = 493) yielded a positive (theory-consistent) but non-significant effect size estimate, $r = 0.082$ (95% CI = -0.065, 0.224), $p = 0.274$ (Fig. 1 for forest plot). A moderate level of heterogeneity was observed, $Q(8) = 18.251$, $p = 0.019$, $\tau^2 = 0.026$, $I^2 = 56.30\%$ (I^2: 25% = low; 50% = moderate; 75% = high [Higgins et al., 2003]), yet moderator analysis showed no statistically significant effect of sex, $Q(1) = 0.250$, $p = 0.617$. The 'leave one out' procedure, which recalculates the meta-analytic estimates whilst omitting each sample consecutively, showed that these findings only changed noticeably if the female subsample from the study by Auyeung et al. (2009) was removed. In this case, the effect size estimate became noticeably smaller, $r = 0.032$ (95% CI = -0.079, 0.143), $p = 0.569$, and significant heterogeneity was no longer observed, $Q(7) = 5.166$, $p = 0.640$, $\tau^2 = 0.003$, $I^2 = 12.49\%$.

Egger's regression did not detect the presence of publication bias, $z = -1.326$, $p = 0.185$. However, it has been suggested that this test lacks sensitivity (Higgins et al., 2003), and indeed three missing studies were estimated via the trim and fill procedure (Duval & Tweedie, 2000). When imputed, the resulting model returned a significant positive correlation, $r = 0.166$ (95% CI = 0.034, 0.293), $p = 0.014$, again with a moderate level of between-sample heterogeneity, $Q(11) = 27.649$, $p = 0.004$, $\tau^2 = 0.032$, $I^2 = 61.57\%$ (Fig. 2 for contour-enhanced funnel plot).

Power calculations using G*Power 3.1 (Faul et al., 2009) were conducted to determine the sample size that would be required to observe a statistically significant effect for a two-tailed Pearson's correlation with α set at $p < 0.05$ and 80% power. Based on the uncorrected effect size estimate

Table 1 Summary of previously reported correlations between amniotic sex hormone concentrations and measures of children's sexually differentiated play behaviours

Cohort location	Study	Journal	Inclusion/exclusion criteria	Maternal age (years)	Gestational age at amniocentesis (weeks)	Child age at follow-up	Hormone assay	Play measure	Analysis	n	r	p
Toronto, Canada	Grimshaw et al. (1995)	Brain and Cognition	Mothers were healthy and eligible for amniocentesis due to age (> 35 years at time of child's birth)	~ 35 years	14–20 (corrected as best estimate of T level at median gestational age (16 weeks)	7 years (within two weeks of 7th birthday)	RIA	Spatial play items from Child Games Inventory	Boys / spatial play	13	-	ns
									Girls / spatial play	12	-	ns
Cambridge, UK	Knickmeyer et al., (2005b)	Developmental Psychology	Mothers referred for amniocentesis due to age (25%) or high result on triple test (60%). Children with medical records indicating illness at birth were not included	Boys' mothers (n = 31): 25–42 (M = 35.6, SD = 5.01) Girls' mothers (n = 22): 23–40 (M = 34.2, SD = 4.63)	Boys (n = 31): 14–20 (M = 16.8, SD = 1.79) Girls (n = 22): 14–18 (M = 16.8, SD = 0.97)	4.75–5.8 years	RIA	Children's Play Questionnaire (maternal report)	Boys / masculine play	31	0.00	0.98
									Boys / feminine play	31	-0.01	0.94
									Girls / masculine play	21	-0.19	0.40
									Girls / feminine play	21	0.36	0.10
	Auyeung et al. (2009)	Psychological Science	Mothers referred for routine amniocentesis who gave birth to healthy infants	Boys' mothers (n = 100): 23.38–44.14 (M = 34.90, SD = 4.78) Girls' mothers (n = 91): 22.81–46.42 (M = 35.31, SD = 4.10)	Boys (n = 78): 13.0–20.0 (M = 16.32, SD = 1.5) Girls (n = 67): 11.0–19.0 (M = 16.31, SD = 2.32)	Boys (n = 110): 6.57–10.29 years (M = 8.70, SD = 0.93) Girls (n = 99): 6.35–10.27 years (M = 8.44, SD = 1.02)	RIA (Count a Coat)	PSAI	Boys / PSAI	112	0.20	< 0.05
									Girls / PSAI	100	0.42	< 0.001
London, UK	Constantinescu (2009)	MPhil thesis	Mothers at increased risk of having a child with Down's syndrome (only those with a healthy foetus were retained)	28.17–44.08 (n = 66, M = 37.67, SD = 4.00) (age at birth)	15–37 (M = 17, SD = 3.7)	3.83–5.92 years (n = 66, M = 4.50, SD = 0.62)	RIA	PSAI	Boys / PSAI	29	-0.106	0.584
									Girls / PSAI	22	-0.182	0.418
	Spencer et al. (2021)	Hormones and Behavior	Mothers at increased risk of having a child with Down's syndrome (only those with a healthy foetus were retained) (as reported in Constantinescu, 2009)	M = 37.3 (SD = 4.05) (age at birth) (as reported in Kung et al. 2016)	15–25 weeks (M = 16.93, SD = 2.01)	Boys: M = 4.31 years (SE = 0.06) Girls: M = 4.23 years (SE = 0.04)	RIA	PSAI	Boys / PSAI	44	-0.03	ns
								Gender typing of favourite toys (child interview)	Boys / toy preference (child)	44	-0.09	ns
								Gender typing of favourite toys (parent interview)	Boys / toy preference (parent)	44	0.31	< 0.05
								Sex of playmate preference	Boys / playmate preference	44	-0.15	ns
								Proportion of time with male-typical toys (8-min observation)	Boys / male-typical toy (observation)	44	0.04	ns
								Proportion of time with female-typical toys (8-min observation)	Boys / female-typical toy (observation)	44	-0.05	ns
									Girls / PSAI	48	0.003	ns
									Girls / toy preference (child)	48	0.14	ns
									Girls / toy preference (parent)	48	-0.05	ns
									Girls / playmate preference	48	0.002	ns
									Girls / male-typical toy (observation)	48	-0.24	ns
									Girls / female-typical toy (observation)	48	0.29	< 0.05

Table 1 (continued)

Cohort location	Study	Journal	Inclusion/exclusion criteria	Maternal age (years)	Gestational age at amniocentesis (weeks)	Child age at follow-up	Hormone assay	Play measure	Analysis	n	r	p
Utrecht, Netherlands	van de Beek et al. (2009)	*Archives of Sexual Behavior*	Mothers referred for amniocentesis due to being 36 years or older (96%), medical history (e.g., a previous child with Down's syndrome or radiation cancer treatment) (3.3%) or high result on triple test (0.7%) who gave birth to healthy singleton infants	28–45 ($M=37.5$, $SD=2.1$) (at amniocentesis)	15.3–18.0 ($M=16.3$, $SD=0.46$)	53.6–60.6 weeks ($M=56.4$, $SD=1.2$)	RIA (log transformed; standardized residuals used in analyses)	7-min structured toy play session	Boys / masculine play	61	0.11	ns
									Boys / feminine play	61	0.01	ns
									Girls / masculine play	61	0.00	ns
									Girls / feminine play	61	0.03	ns
	Beking (2018)	PhD thesis	Same as van de Beek et al. (2009); two boys excluded (one for playing less than 10% of the time; one for playing with too many toys simultaneously)	36–42 (p. 62) (due to advanced maternal age) and n=7 aged 29–35 due to risk factors in their medical history (p. 122)	15.3–18.2 ($M=16.4$, $SD=0.48$)	13 months	RIA	7-min structured toy play session	Boys / masculine/feminine play	134 (M+F)	-	ns
									Girls / masculine/feminine play		-	ns
			Data for one girl lost due to technical problems			2.5 years	RIA	7-min structured toy play session	Boys / masculine/feminine play	89 (M+F)	-	ns
									Girls / masculine/feminine play		-	ns
			3 boys and 1 girl excluded for playing less than 50% of the time			6.5 years	RIA	7-min structured toy play session	Boys / masculine/feminine play	66 (M+F)	-	ns
									Girls / masculine/feminine play		-	0.041
Düsseldorf, Germany	Körner (2018)	PhD thesis	Mothers with no identified prenatal abnormalities and an uncomplicated pregnancy; children with APGAR score of at least 9	22–48 ($M=38.13$, $SD=3.50$) (at delivery)	14–18 ($M=14.84$, $SD=0.91$)	40 months	LC/MS	PSAI (German version)	Boys / PSAI	42	-0.12	0.461

Note: LC/MS = liquid chromatography and tandem mass spectrometry; PSAI = Pre-School Activities Inventory; RIA = radioimmunoassay; T = testosterone

($r = 0.082$), n = 1,165 participants would be required; if using the effect size estimate corrected for hypothesised publication bias ($r = 0.166$), n = 282 would be required. However, if basing the calculation on the effect size produced when the female subsample from Auyeung et al. (2009) was omitted ($r = 0.032$), a much larger sample (n = 7,662) would be necessary.

Discussion

The current paper presents a meta-analysis to test the hypothesis that second trimester amniotic testosterone concentrations correlate positively with male-typical play preferences and negatively with female-typical play preferences during infancy/childhood. Although amniocentesis is rarely utilised within the behavioural sciences, we were able to identify nine samples that met our inclusion criteria (five male, four female). These samples together represent five independent cohorts: one each from Canada, Germany, and the Netherlands, and two from the UK. Although the initial effect size estimate returned by the meta-analysis was in the theory-consistent direction, it was not statistically significant ($r = 0.082$, $p = 0.274$). However, when employing Duval and Tweedie's (2000) trim and fill procedure to adjust for inferred publication bias, the effect size estimate doubled in magnitude and became statistically significant ($r = 0.166$, $p = 0.014$). The reason(s) for this is/are unclear, though our findings as a whole appear to be in line with the idea that individual differences in foetal testosterone concentration during the second trimester relate to sexually differentiated play preferences in infancy/childhood. Interestingly, this effect appears to be small in magnitude, and there was no moderating effect of sex, which suggests that prenatal testosterone exerts a similar influence on the subsequent play preferences of females as it does for males.

Our findings are generally consistent with the broader literature examining associations between prenatal testosterone and sexually differentiated play preferences, as well as with the notion that it is during the second trimester of pregnancy that testosterone exerts its largest influence on brain development (Baron-Cohen et al., 2004; Hines et al., 2002). However, it should be noted that moderate between-sample heterogeneity in the effect size estimates was observed. Although this observation suggests that systematic differences between studies likely contribute to variation in their outcome, we only conducted a moderation analysis for sex (male/female). We ran the analysis because of the obvious importance of this variable in the current context, and because sex-stratified data were available for each of the identified studies. We did not conduct moderation analyses to explore other potential sources of heterogeneity because the reliability is known to be low when relatively few samples are included

(Borenstein et al., 2009). Of note, it would have been particularly useful to consider different types of childhood play measures separately because toy preferences appear to show a particularly large sex difference, and because the magnitude of the sex difference is known to vary as a function of the method used as well as the types of toys (Hines, 2013a, b). Thus, the current study was unable to identify specific features of play which may be driving the observed effect.

Interestingly, the level of heterogeneity decreased markedly (as did the overall effect size estimate) if the female subsample from Auyeung et al. (2009) was omitted from analysis. This sample notably represented not only the largest effect size ($r = 0.42$) within the meta-analysis, but also the second largest sample size (n = 100). It therefore does not typify the undue influence sometimes exerted by small studies that report inflated effect size estimates. Furthermore, as an effect size from one of the largest samples included is often a good indicator of the overall outcome of a meta-analysis (Peters et al., 2007), this finding is not easy to interpret.

Although prenatal testosterone appears to play a role in the development of sexually differentiated play preferences, the small magnitude of the correlation (Cohen, 1988) derived from the meta-analysis means that most of the variance remains unexplained. Some of this variance undoubtedly reflects measurement error (both for amniotic testosterone as well as for infant/child play). A further consideration is the influence of additional environmental effects on play preferences, such as those exerted by parents, teachers, and peers. It seems most likely that a complex interaction occurs by which biological factors (e.g., differences in prenatal testosterone exposure) modify later environmental influences (e.g., parenting style) and that these environmental influences also modify the effects of those initial biological predispositions (Hines, 2013b; Udry, 2003). For instance, a child with a biological tendency toward sex-typical play preferences may elicit different responses from parents than would a child with biological tendencies toward sex-atypical play; these differential responses might then interact with the underlying biological predispositions to further modify the child's behaviour. Ultimately, play behaviours are invariably the result of an intricate interplay between biological and social factors which act throughout development, starting very early in life.

It is crucial to reemphasise that a statistically significant meta-analytic effect size estimate was only observed in the current study after we implemented the trim and fill procedure to correct for hypothesised publication bias. It is also important to note that the trim and fill procedure is based on funnel plot asymmetry, and that this phenomenon can occur for reasons other than publication bias, such as heterogeneity in study quality, language of publication and adequacy of the statistical analysis, as well as chance (see Egger et al., 1997). Furthermore, the trim and fill procedure was initially

Table 2 Results of two-level and three-level meta-analyses of the association between amniotic testosterone concentration and sexually differentiated play preferences

Model	Meta-analytic effect size estimate			Heterogeneity				
	r (95% CI)	SE	p	Q	df	p	τ^2	I^2
Two-level	0.082 (-0.065, 0.224)	0.075	0.274	18.251	8	0.019	0.026	56.30%
Three-level	0.061 (-0.151, 0.266)	0.092	0.529					

intended as a form of sensitivity analysis (Duval & Tweedie, 2000), and a detailed simulation study by Peters et al. (2007) generally upholds this suggestion. However, although Peters et al. (2007) noted that performance of the method is 'not ideal', they also stated that "when publication bias is present the trim and fill method can give estimates that are less biased than the usual meta-analysis models" (p. 4544). They also noted that when there is substantial between-study heterogeneity in addition to evidence of publication bias, it may be appropriate to "give more weight to conclusions based on findings from the random-random effects trim and fill model" than the unadjusted model (p. 4557). Caution is, however, required when interpreting our findings because

the number of samples included in the meta-analysis was small and so the imputation of three hypothesised studies will have exerted a relatively large influence on the outcome. Furthermore, publication bias might seem unlikely when considering that all peer-reviewed findings except those of Auyeung et al. (2009) were non-significant. However, it is notable that one study was presented solely within a PhD thesis (Körner, 2018), an earlier analysis of the Spencer et al. (2021) cohort appeared in an MPhil thesis a decade earlier (Constantinescu, 2009), and the longitudinal follow-ups of van de Beek et al. (2009) are in a PhD thesis (Beking, 2018) but not in a scientific journal article. Nevertheless, as each of these cohorts is represented within the meta-analysis, it

Fig. 1 Forest plot displaying correlation between amniotic testosterone concentration and children's sexually differentiated play preferences for each sample included within the meta-analysis. Note. Square boxes and lines represent effect sizes and 95% confidence intervals, respectively, and the diamond indicates the overall meta-analytic effect size estimate. F = female sample; M = male sample
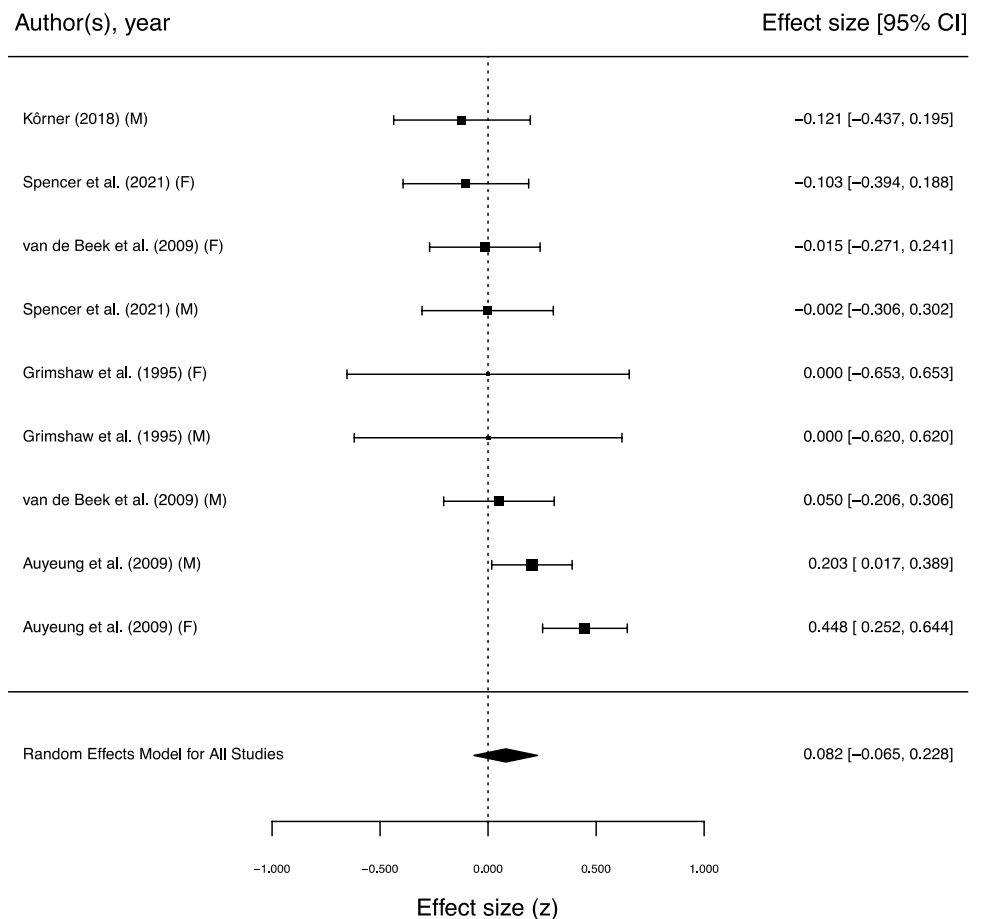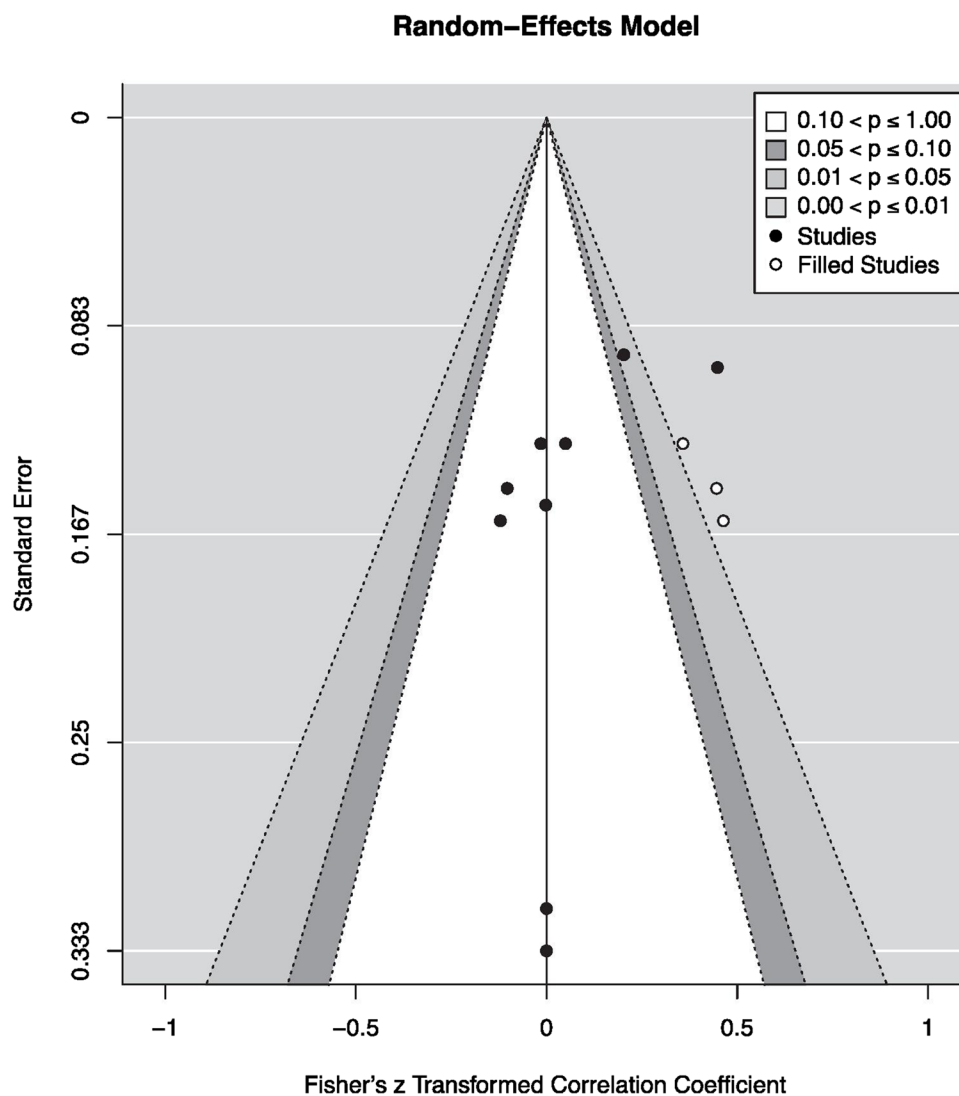
Fig. 2 Contour-enhanced funnel plot for meta-analysis of amniotic testosterone concentration and children's sexually differentiated play preferences. Note. The filled circles represent individual samples, and the unfilled circles represent hypothesised missing studies

is intriguing that hypothesised missing studies were still detected. This outcome could imply that publication bias really is an issue within this literature, mirroring suggestions about research on CAH (Collaer & Hines, 2020; Hampson, 2016; Richards et al., 2020). Regardless, considering the substantial difficulties associated with setting up amniotic fluid studies (e.g., ethics, finance, technical expertise, time-constraints, effective collaboration with clinical staff already working full-time hours etc.), we encourage researchers to report the findings for all outcomes irrespective of their direction or level of statistical significance.

In the current study, we took the same approach as a recent review of behavioural sex differences (Xiong & Scott, 2020) by focusing on the specific paradigm of amniotic fluid studies. Although advantageous because this methodology has been hailed as the most informative in terms of prenatal hormone-behaviour associations (Baron-Cohen et al., 2004; van de Beek et al., 2004), it remains questionable whether

testosterone assayed via a stressful medical procedure (Ventura et al., 2012) conducted at a single and variable timepoint in samples which may not be entirely generalisable, are a valid approximation of those present throughout gestation. Indeed, the only study to correlate amniotic testosterone with that of the actual foetal circulation (Rodeck et al., 1985) reported no statistically significant relationship. Therefore, further exploration of the validity of this method is arguably warranted. However, as amniocentesis is now rarely performed (Akolekar et al., 2015), researchers may also need to look towards more recently developed techniques of a less invasive nature.

Our analysis is limited in that it does not consider the possible influence of early postnatal testosterone exposure. Males notably experience a second surge in testosterone, sometimes referred to as 'mini-puberty' (for a review, see Lanciotti et al., 2018), that is produced largely from the testes during the first few postnatal months (Forest, 1990;

Forest et al., 1974; Quigley, 2002). Because brain development continues into the second year of postnatal life, this testosterone surge could plausibly influence the development of a male-typical behavioural phenotype (Hines, 2011; Hines et al., 2016). Although testosterone measured from second trimester amniotic fluid and early postnatal saliva samples appear to be uncorrelated (Auyeung et al., 2012), and associations between CAH status and male-typical play behaviour may be more consistent with prenatal than postnatal androgenic influence (Berenbaum et al., 2000), some studies have examined correlations between mini-pubertal testosterone and childhood play preferences. Notably, Lamminmäki et al. (2012) reported theory-consistent associations between infant play preferences and testosterone measured from urinary samples obtained at repeated intervals throughout the first six months of postnatal life. A strength of their study is that sexually differentiated behaviour was assessed several months after the testosterone samples were obtained. This feature is important because it would allow sufficient time for the testosterone to have exerted organisational influences on the brain and behaviour, meaning that the observations made could not be attributable to transient (i.e., activational) effects of the hormone. Findings from studies that utilised saliva samples, on the other hand, have generally been equivocal (Alexander & Saenz, 2012; Alexander et al., 2009). It is therefore relevant to note that these, and some other studies (Alexander & Saenz, 2011, 2012; Auyeung et al., 2012; Corpuz, 2021) that obtained saliva samples at approximately 3–6 months did not observe a significant sex difference for testosterone. It therefore seems plausible that the mini-pubertal testosterone peak is diminished by this point and so may not be detectable when measured from saliva samples. This notion is consistent with Huhtaniemi et al. (1986), who reported that salivary testosterone concentrations in 22 male infants were highest at 2–10 days of postnatal life. We therefore suggest that it may be fruitful for future research to focus on testosterone measured soon after birth rather than that measured several months later.

When interpreting behavioural sex differences in the context of normal variations in prenatal and early postnatal testosterone concentrations, it is important to consider other variables which may have contributed to an observed relationship. The sexual differentiation of the brain and behaviour, compared to that of the sexual structures, is complex and encompasses a broader timeframe and various environmental and biological events (Jordan-Young, 2010). Also, a single sample of testosterone will not provide a completely accurate measure of the hormonal milieu experienced by the foetus or young infant. However, ethical constraints preclude the opportunity to study hormones more directly during times of early brain development in humans. Another factor to consider is that, although the presence of testosterone has been shown to influence sexually differentiated

behaviours, higher testosterone does not inevitably promote a male-typical phenotype for all behaviours (Hines, 2004). Indeed, certain behaviours could be more susceptible to influence from the prenatal and/or early postnatal hormonal environment than others. Also, different sexually differentiated behaviours may have different critical periods of development. This idea is consistent with the observation that exposure to testosterone during early and late periods of prenatal development has differential effects on subsequently measured male-typical behaviours in female rhesus monkey (Goy et al., 1988) and rodent models (Rhees et al., 1997).

Conclusions

The current report presents a meta-analysis of studies linking amniotic testosterone with sexually differentiated human play preferences in infancy/childhood. Although the initial meta-analysis returned a null result, a statistically significant positive correlation emerged once hypothesised publication bias was controlled for. However, one specific sample appeared to exert a particularly large influence on the outcome of our meta-analysis, and questions regarding publication bias within this field should encourage researchers with relevant data to publish their results, regardless of direction of effect or degree of statistical significance. We also suggest that the mini-pubertal testosterone surge warrants further investigation, and that researchers in this field should consider focusing their attention on samples obtained soon after birth, as sex differences do not appear to be reliably detected several months later.

Declarations

References

Abramovich, D. R. (1974). Human sexual differentiation—in Utero Influences. *BJOG: an International Journal of Obstetrics and Gynaecology, 81*(6), 448–453. https://doi.org/10.1111/j.1471-0528.1974.tb00494.x.

Akolekar, R., Beta, J., Picciarelli, G., Ogilvie, C., & D'Antonio, F. (2015). Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: A systematic review and meta-analysis. *Ultrasound in Obstetrics & Gynecology, 45*(1), 16–26. https://doi.org/10.1002/uog.14636.

Alexander, G. M., & Saenz, J. (2011). Postnatal testosterone levels and temperament in early infancy. *Archives of Sexual Behavior, 40*(6), 1287–1292. https://doi.org/10.1007/s10508-010-9701-5.

Alexander, G. M., & Saenz, J. (2012). Early androgens, activity levels and toy choices of children in the second year of life. *Hormones and Behavior, 62*(4), 500–504. https://doi.org/10.1016/j.yhbeh.2012.08.008.

Alexander, G. M., Wilcox, T., & Farmer, M. E. (2009). Hormone-behavior associations in early infancy. *Hormones and Behavior, 56*(5), 498–502. https://doi.org/10.1016/j.yhbeh.2009.08.003.

Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R. C., Taylor, K., Hackett, G., & Hines, M. (2009). Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. *Psychological Science, 20*(2), 144–148. https://doi.org/10.1111/j.1467-9280.2009.02279.x.

Auyeung, B., Ahluwalia, J., Thomson, L., Taylor, K., Hackett, G., O'Donnell, K. J., & Baron-Cohen, S. (2012). Prenatal versus postnatal sex steroid hormone effects on autistic traits in children at 18 to 24 months of age. *Molecular Autism, 3*(1), 17. https://doi.org/10.1186/2040-2392-3-17.

Baron-Cohen, S., Lutchmaya, S., & Knickmeyer, R. C. (2004). *Prenatal testosterone in mind*. MIT Press.

Baron-Cohen, S., Auyeung, B., Nørgaard-Pedersen, B., Hougaard, D. M., Abdallah, M. W., Melgaard, L., Cohen, A. S., Chakrabarti, B., Ruta, L., & Lombardo, M. V. (2015). Elevated fetal steroidogenic activity in autism. *Molecular Psychiatry, 20*(3), 369–376. https://doi.org/10.1038/mp.2014.48.

Beking, T. (2018). *Two sides to every story: Sex hormones, brain lateralization and gender development*. University of Groningen.

Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science, 3*(3), 203–206. https://doi.org/10.1111/j.1467-9280.1992.tb00028.x.

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*(1), 31–42. https://doi.org/10.1037/0012-1649.31.1.31.

Berenbaum, S. A., Duck, S. C., & Bryk, K. (2000). Behavioral effects of prenatal versus postnatal androgen excess in children with 21-hydroxylase-deficient congenital adrenal hyperplasia. *Journal of Clinical Endocrinology & Metabolism, 85*(2), 727–733. https://doi.org/10.1210/jc.85.2.727.

Blakemore, J. E. O., Berenbaum, S. A., & Liben, L. S. (2009). *Gender development*. Psychology Press.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386.

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

Cohen-Bendahan, C. C. C., van de Beek, C., & Berenbaum, S. A. (2005). Prenatal sex hormone effects on child and adult sex-typed behavior: Methods and findings. *Neuroscience & Biobehavioral Reviews, 29*(2), 353–384. https://doi.org/10.1016/j.neubiorev.2004.11.004.

Collaer, M. L., & Hines, M. (2020). No evidence for enhancement of spatial ability with elevated prenatal androgen exposure in congenital adrenal hyperplasia: A meta-analysis. *Archives of Sexual Behavior, 49*(2), 395–411. https://doi.org/10.1007/s10508-020-01645-7.

Constantinescu, M. (2009). *Are finger ratios a useful measure of androgenic influences on sexual differentiation?* University of Cambridge.

Corpuz, R. (2021). The role of maternal environment on calibrating "Mini Puberty" in early infant development. *Developmental Psychobiology, 63*(4), 800–807. https://doi.org/10.1002/dev.22033.

Davis, J. T. M., & Hines, M. (2020). How large are gender differences in toy preferences? A systematic review and meta-analysis of toy preference research. *Archives of Sexual Behavior, 49*(2), 373–394. https://doi.org/10.1007/s10508-019-01624-7.

DiPietro, J. A. (1981). Rough and tumble play: A function of gender. *Developmental Psychology, 17*(1), 50–58. https://doi.org/10.1037/0012-1649.17.1.50.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629.

Fagen, R. (1981). *Animal play behavior*. Oxford University Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Forest, M. G. (1990). Pituitary gonadotropin and sex steroid secretion during the first two years of life. In M. M. Grumbach, P. C. Sizonenko, & M. L. Aubert (Eds.), *Control of the onset of puberty* (pp. 451–478). Williams & Wilkins.

Forest, M. G., Sizonenko, P. C., Cathiard, A. M., & Bertrand, J. (1974). Hypophyso-gonadal function in humans during the first year of life: I. Evidence for testicular activity in early infancy. *Journal of Clinical Investigation, 53*(3), 819–828. https://doi.org/10.1172/jci107621.

Goy, R. W., Bercovitch, F. B., & McBrair, M. C. (1988). Behavioral masculinization is independent of genital masculinization in prenatally androgenized female rhesus macaques. *Hormones and Behavior, 22*(4), 552–571. https://doi.org/10.1016/0018-506X(88)90058-X.

Grimshaw, G. M., Sitarenios, G., & Finegan, J. A. K. (1995). Mental rotation at 7 years - relations with prenatal testosterone levels and spatial play experiences. *Brain and Cognition, 29*(1), 85–100. https://doi.org/10.1006/brcg.1995.1269.

Hampson, E. (2016). The development of hand preference and dichotic language lateralization in males and females with congenital adrenal hyperplasia. *Laterality, 21*(4–6), 415–432. https://doi.org/10.1080/1357650X.2015.1102924.

Harris, J. A., Vernon, P. A., & Boomsma, D. I. (1998). The heritability of testosterone: A study of dutch adolescent twins and their parents. *Behavior Genetics, 28*(3), 165–171. https://doi.org/10.1023/A:1021466929053.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557.

Hines, M. (2004). *Brain gender*. Oxford University Press.

Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends in Cognitive Sciences, 14*(10), 448–456. https://doi.org/10.1016/j.tics.2010.07.005.

Hines, M. (2011). Gender development and the human brain. *Annual Review of Neuroscience, 34*(1), 69–88. https://doi.org/10.1146/annurev-neuro-061010-113654.

Hines, M. (2013a). Sex and sex differences. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology (Vol. 1): Body and mind* (pp. 164–201). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199958450.013.0007.

Hines, M. (2013b). Gonadal hormone influences on human neurobehavioral development: Outcomes and mechanisms. In D. W. Pfaff & Y. Christen (Eds.), *Multiple origins of sex differences in brain: Neuroendocrine functions and their pathologies* (pp. 59–69). Springer. https://doi.org/10.1007/978-3-642-33721-5_5.

Hines, M. (2015). Gendered development. In R. M. Lerner & M. E. Lamb (Eds.), *Handbook of child psychology and developmental science: Volume 3. Socioemotional processes* (pp. 842–887). Wiley. https://doi.org/10.1002/9781118963418.childpsy320.

Hines, M., & Kaufman, F. R. (1994). Androgen and the development of human sex-typical behavior: Rough-and-tumble play and sex of preferred playmates in children with congenital adrenal hyperplasia (CAH). *Child Development, 65*(4), 1042–1053. https://doi.org/10.1111/j.1467-8624.1994.tb00801.x.

Hines, M., Golombok, S., Rust, J., Johnston, K. J., Golding, J., & Avon Longitudinal Study of Parents and Children Study Team. (2002). Testosterone during pregnancy and gender role behavior of preschool children: A longitudinal, population study. *Child Development, 73*(6), 1678–1687. https://doi.org/10.1111/1467-8624.00498.

Hines, M., Ahmed, S. F., & Hughes, I. A. (2003). Psychological outcomes and gender-related development in complete androgen insensitivity syndrome. *Archives of Sexual Behavior, 32*(2), 93–101. https://doi.org/10.1023/a:1022492106974.

Hines, M., Brook, C., & Conway, G. S. (2004). Androgen and psychosexual development: Core gender identity, sexual orientation, and recalled childhood gender role behavior in women and men with congenital adrenal hyperplasia (CAH). *Journal of Sex Research, 41*(1), 75–81. https://doi.org/10.1080/00224490409552215.

Hines, M., Spencer, D., Kung, K. T. F., Browne, W. V., Constantinescu, M., & Noorderhaven, R. M. (2016). The early postnatal period, mini-puberty, provides a window on the role of testosterone in human neurobehavioural development. *Current Opinion in Neurobiology, 38*, 69–73. https://doi.org/10.1016/j.conb.2016.02.008.

Hönekopp, J., & Watson, S. (2011). Meta-analysis of the relationship between digit-ratio 2D:4D and aggression. *Personality and Individual Differences, 51*(4), 381–386. https://doi.org/10.1016/j.paid.2010.05.003.

Huhtaniemi, I., Dunkel, L., & Perheentupa, J. (1986). Transient increase in postnatal testicular activity is not revealed by longitudinal measurements of salivary testosterone. *Pediatric Research, 20*(12), 1324–1327. https://doi.org/10.1203/00006450-19861200-00028.

Jordan-Young, R. M. (2010). *Brainstorm: The flaws in the science of sex differences*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9w21.

Knickmeyer, R. C., Baron-Cohen, S., Raggatt, P., & Taylor, K. (2005a). Foetal testosterone, social relationships, and restricted interests in children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 46*(2), 198–210. https://doi.org/10.1111/j.1469-7610.2004.00349.x.

Knickmeyer, R. C., Wheelwright, S., Taylor, K., Raggatt, P., Hackett, G., & Baron-Cohen, S. (2005b). Gender-typed play and amniotic testosterone. *Developmental Psychology, 41*(3), 517–528. https://doi.org/10.1037/0012-1649.41.3.517.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2*(1), 61–76. https://doi.org/10.1002/jrsm.35.

Körner, L. M. (2018). *Geschlechterunterschiede im spielverhalten und der selbstkontrolle bei 40 monate alten kindern und der zusammenhang mit pränatalen sexualhormonen* [Heinrich-Heine-Universität Düsseldorf]. https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=46130.

Kung, K. T. F., Spencer, D., Pasterski, V. L., Neufeld, S. A. S., Glover, V., Connor, T. G. O., Hindmarsh, P. C., Hughes, I. A., Acerini, C. L., & Hines, M. (2016). No relationship between prenatal androgen exposure and autistic traits: Convergent evidence from studies of children with congenital adrenal hyperplasia and of amniotic testosterone concentrations in typically developing children. *Journal of Child Psychology and Psychiatry, 57*(12), 1455–1462. https://doi.org/10.1111/jcpp.12602.

LaFreniere, P. (2011). Evolutionary functions of social play: Life histories, sex differences, and emotion regulation. *American Journal of Play, 3*(4), 464–488.

Lamminmäki, A., Hines, M., Kuiri-Hänninen, T., Kilpeläinen, L., Dunkel, L., & Sankilampi, U. (2012). Testosterone measured in infancy predicts subsequent sex-typed behavior in boys and in girls. *Hormones and Behavior, 61*(4), 611–616. https://doi.org/10.1016/j.yhbeh.2012.02.013.

Lanciotti, L., Cofini, M., Leonardi, A., Penta, L., & Esposito, S. (2018). Up-to-date review about minipuberty and overview on hypothalamic-pituitary-gonadal axis activation in fetal and neonatal life. *Frontiers in Endocrinology, 9*, 410. https://doi.org/10.3389/fendo.2018.00410.

Maccoby, E. E. (1988). Gender as a social category. *Developmental Psychology, 24*(6), 755–765. https://doi.org/10.1037/0012-1649.24.6.755.

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. The Belknap Press of Harvard University Press.

Maccoby, E. E., & Jacklin, C. N. (1987). Gender segregation in children. In H. W. Reece (Ed.), *Advances in Child Development and Behavior* (Vol. 20, pp. 239–287). https://doi.org/10.1016/S0065-2407(08)60404-8.

Martin, C. R. (1985). *Endocrine physiology*. Oxford University Press.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology, 37*(3), 431–446. https://doi.org/10.1037/0012-1649.37.3.431.

Merke, D. P., & Bornstein, S. R. (2005). Congenital adrenal hyperplasia. *Lancet, 365*(9477), 2125–2136. https://doi.org/10.1016/S0140-6736(05)66736-0.

Meyer-Bahlburg, H. F. L., Dolezal, C., Baker, S. W., Carlson, A. D., Obeid, Jihad, S., & New, M. I. (2004). Prenatal androgenization affects gender-related behavior but not gender identity in 5-12-year-old girls with congenital adrenal hyperplasia. *Archives of Sexual Behavior, 33*(2), 97–104.

Nagamani, M., McDonough, P. G., Ellegood, J. O., & Mahesh, V. B. (1979). Maternal and amniotic fluid steroids throughout human pregnancy. *American Journal of Obstetrics and Gynecology, 134*(6), 674–680. https://doi.org/10.1016/0002-9378(79)90649-5.

Nass, R., & Baker, S. (1991). Androgen effects on cognition: Congenital adrenal hyperplasia. *Psychoneuroendocrinology, 16*(1–3), 189–201. https://doi.org/10.1016/0306-4530(91)90078-8.

Nassar, G. N., & Leslie, S. W. (2021). Physiology, Testosterone. In *StatPearls*. StatPearls Publishing LLC. https://www.ncbi.nlm.nih.gov/books/NBK526128/.

Nordenström, A., Servin, A., Bohlin, G., Larsson, A., & Wedell, A. (2002). Sex-typed toy play behavior correlates with the degree of prenatal androgen exposure assessed by CYP21 genotype in girls with congenital adrenal hyperplasia. *Journal of Clinical Endocrinology & Metabolism, 87*(11), 5119–5124. https://doi.org/10.1210/jc.2001-011531.

Open Science Collaboration, Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716.

Pasterski, V. L. (2008). Disorders of sex development and atypical sex differentiation. In D. L. Rowland & L. Incrocci (Eds.), *Handbook of sexual and gender identity disorders* (pp. 354–375). Wiley.

Pasterski, V. L., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2005). Prenatal hormones and postnatal socialization by parents as determinants of male-typical toy play in girls with congenital adrenal hyperplasia. *Child Development, 76*(1), 264–278. https://doi.org/10.1111/j.1467-8624.2005.00843.x.

Pasterski, V. L., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2011). Prenatal hormones and childhood sex segregation: Playmate and play style preferences in girls with congenital adrenal hyperplasia. *Hormones and Behavior, 59*(4), 549–555. https://doi.org/10.1016/j.yhbeh.2011.02.007.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26*(25), 4544–4562. https://doi.org/10.1002/sim.2889.

Quigley, C. A. (2002). Editorial: The postnatal gonadotropin and sex steroid surge–Insights from the androgen insensitivity syndrome. *Journal of Clinical Endocrinology & Metabolism, 87*(1), 24–28. https://doi.org/10.1210/jc.87.1.24.

Rhees, R. W., Kirk, B. A., Sephton, S., & Lephart, E. D. (1997). Effects of prenatal testosterone on sexual behavior, reproductive morphology and LH secretion in the female rat. *Developmental Neuroscience, 19*(5), 430–437. https://doi.org/10.1159/000111240.

Richards, G., Browne, W. V., Aydin, E., Constantinescu, M., Nave, G., Kim, M. S., & Watson, S. J. (2020). Digit ratio (2D:4D) and congenital adrenal hyperplasia (CAH): Systematic literature review and meta-analysis. *Hormones and Behavior, 126*, 104867. https://doi.org/10.1016/j.yhbeh.2020.104867.

Rodeck, C. H., Gill, D., Rosenberg, D. A., & Collins, W. P. (1985). Testosterone levels in midtrimester maternal and fetal plasma and amniotic fluid. *Prenatal Diagnosis, 5*(3), 175–181. https://doi.org/10.1002/pd.1970050303.

Sato, T., Matsumoto, T., Kawano, H., Watanabe, T., Uematsu, Y., Sekine, K., Fukuda, T., Aihara, K., Krust, A., Yamada, T., Nakamichi, Y., Yamamoto, Y., Nakamura, T., Yoshimura, K., Yoshizawa, T., Metzger, D., Chambon, P., & Kato, S. (2004). Brain masculinization requires androgen receptor function. *Proceedings of the National Academy of Sciences of the United States of America, 101*(6), 1673–1678. https://doi.org/10.1073/pnas.0305303101.

Servin, A., Nordenström, A., Larsson, A., & Bohlin, G. (2003). Prenatal androgens and gender-typed behavior: A study of girls with mild and severe forms of congenital adrenal hyperplasia. *Developmental Psychology, 39*(3), 440–450. https://doi.org/10.1037/0012-1649.39.3.440.

Smith, P. K. (1982). Does play matter? Functional and evolutionary aspects of animal and human play. *Behavioral and Brain Sciences, 5*(1), 139–155. https://doi.org/10.1017/S0140525X0001092X.

Smith, P. K., & Connolly, K. J. (1980). *The ecology of preschool behavior*. Cambridge University Press.

Spencer, D., Pasterski, V. L., Neufeld, S. A. S., Glover, V., O'Connor, T. G., Hindmarsh, P. C., Hughes, I. A., Acerini, C. L., & Hines, M. (2021). Prenatal androgen exposure and children's gender-typed behavior and toy and playmate preferences. *Hormones and Behavior, 127*, 104889. https://doi.org/10.1016/j.yhbeh.2020.104889.

Thorne, B. (2001). Girls and boys together _ but mostly apart: Gender arrangements in elementary schools. In R. Satow (Ed.), *Gender and social life* (pp. 152–166). Pearson.

Todd, B. K., Fischer, R. A., Di Costa, S., Roestorf, A., Harbour, K., Hardiman, P., & Barry, J. A. (2018). Sex differences in children's toy preferences: A systematic review, meta-regression, and meta-analysis. *Infant and Child Development, 27*(2), e2064. https://doi.org/10.1002/icd.2064.

Udry, J. R. (2003). Putting prenatal effects on sex-dimorphic behavior in perspective: An absolutely complete theory. *Epidemiology, 14*(2), 135–136. https://doi.org/10.1097/01.ede.0000054362.61254.b5.

Udry, J. R., Morris, N. M., & Kovenock, J. (1995). Androgen effects on women's gendered behaviour. *Journal of Biosocial Science, 27*(3), 359–368. https://doi.org/10.1017/S0021932000022884.

van de Beek, C., Thijssen, J. H. H., Cohen-Kettenis, P. T., van Goozen, S. H. M., & Buitelaar, J. K. (2004). Relationships between sex hormones assessed in amniotic fluid, and maternal and umbilical cord serum: What is the best source of information to investigate the effects of fetal hormonal exposure? *Hormones and Behavior, 46*(5), 663–669. https://doi.org/10.1016/j.yhbeh.2004.06.010.

van de Beek, C., van Goozen, S. H. M., Buitelaar, J. K., & Cohen-Kettenis, P. T. (2009). Prenatal sex hormones (maternal and amniotic fluid) and gender-related play behavior in 13-month-old infants. *Archives of Sexual Behavior, 38*(1), 6–15. https://doi.org/10.1007/s10508-007-9291-z.

Ventura, T., Gomes, M. C., & Carreira, T. (2012). Cortisol and anxiety response to a relaxing intervention on pregnant women awaiting amniocentesis. *Psychoneuroendocrinology, 37*(1), 148–156. https://doi.org/10.1016/j.psyneuen.2011.05.016.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03.

Warne, G. L., Faiman, C., Reyes, F. I., & Winter, J. S. D. (1977). Studies on human sexual development. V. Concentrations of testosterone, 17-hydroxyprogesterone and progesterone in human amniotic fluid throughout gestation. *Journal of Clinical Endocrinology & Metabolism, 44*(5), 934–938. https://doi.org/10.1210/jcem-44-5-934.

Wong, W. I., Pasterski, V. L., Hindmarsh, P. C., Geffner, M. E., & Hines, M. (2013). Are there parental socialization effects on the sex-typed behavior of individuals with congenital adrenal hyperplasia? *Archives of Sexual Behavior, 42*(3), 381–391. https://doi.org/10.1007/s10508-012-9997-4.

Xiong, H., & Scott, S. (2020). Amniotic testosterone and psychological sex differences: A systematic review of the extreme male brain theory. *Developmental Review, 57*, 100922. https://doi.org/10.1016/j.dr.2020.100922.