Hormones and Behavior 78 (2016) 86–94

Contents lists available at ScienceDirect

Hormones and Behavior

journal homepage: www.elsevier.com/locate/yhbeh

Interpreting digit ratio (2D:4D)–behavior correlations: 2D:4D sex difference, stability, and behavioral correlates and their replicability in young children

Wang I. Wong *[,1], Melissa Hines

Department of Psychology, University of Cambridge, United Kingdom

ARTICLE INFO

Article history:
Received 4 May 2015
Revised 31 October 2015
Accepted 31 October 2015
Available online 2 November 2015

Keywords:
Hormones and behavior
Digit ratio
2D:4D
Sex differences
Early childhood

ABSTRACT

The popularity of using the ratio of the second to the fourth digit (2D:4D) to study influences of early androgen exposure on human behavior relies, in part, on a report that the ratio is sex-dimorphic and stable from age 2 years (Manning etal., 1998). However, subsequent research has rarely replicated this finding. Moreover, although 2D:4D has been correlated with many behaviors, these correlations are often inconsistent. Young children's 2D:4D–behavior correlations may be more consistent than those of older individuals, because young children have experienced fewer postnatal influences. To evaluate the usefulness of 2D:4D as a biomarker of prenatal androgen exposure in studies of 2D:4D–behavior correlations, we assessed its sex difference, temporal stability, and behavioral correlates over a 6- to 8-month period in 126, 2- to 3-year-old children, providing a rare same-sample replicability test. We found a moderate sex difference on both hands and high temporal stability. However, between-sex overlap and within-sex variability were also large. Only 3 of 24 correlations with sex-typed behaviors—scores on the Preschool Activities Inventory (PSAI), preference for a boy-typical toy, preference for a girl-typical toy, were significant and in the predicted direction, all of which involved the PSAI, partially confirming findings from another study. Correlation coefficients were larger for behaviors that showed larger sex differences. But, as in older samples, the overall pattern showed inconsistency across time, sex, and hand. Therefore, although sex-dimorphic and stable, 2D:4D–behavior correlations are no more consistent for young children than for older samples. Theoretical and methodological implications are discussed.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Prenatal androgen exposure influences human behavior and has been linked to variability in many outcomes, including sexual orientation (Meyer-Bahlburg etal., 2008; Zucker etal., 1996), childhood play (Hines etal., 2002), and aggression (Pasterski etal., 2007). However, due to difficulty in measuring prenatal androgen levels directly in typically-developing individuals, much of the research on human prenatal androgen effects has relied on individuals with disorders of sex development (DSDs) which cause unusual androgen exposure. Studies of individuals with DSDs are valuable, but often limited in sample size and generalizability, because clinical samples may differ from typically-developing populations in non-androgen-related variables.

As a result, researchers have used biomarkers of prenatal androgen exposure. One such biomarker is anogenital distance (AGD), the distance between the anus and the genitalia, which is two-fold larger in males than in females at birth (or $d = 3.90$) (Salazar-Martinez etal., 2004) and shows a somewhat smaller difference later in life (Dean and Sharpe, 2013). AGD correlates with prenatal testosterone levels in humans and rats (Callegari etal., 1987; Welsh etal., 2008; Yeh etal., 2002), and AGD at birth has been found to predict gender-relatedbehavior in boys, although not in girls (Pasterski etal., 2015). A second biomarker, the ratio of the second to the fourth digit (2D:4D), is lower in males than in females (left hand $d = .35$; right hand $d = .46$) (see meta-analysis by Hönekopp and Watson, 2010). Although the sex difference in 2D:4D is only small to moderate in size, this biomarker has been widely used in humans. Hundreds of studies have reported that 2D:4D relates to outcomes, including autism (Manning etal., 2001), aggression (Bailey and Hurd, 2005), academic achievement (Brosnan, 2008), and sexual orientation (Grimbos etal., 2010). The use of 2D:4D was also rising as of 2009 (Voracek and Loibl, 2009), with studies typically using 2D:4D as a proxy for prenatal androgen exposure in the study ofhormone–behavior associations. Nevertheless, researchers

* Corresponding author at: Department of Psychology, University of Hong Kong, Hong Kong.
E-mail address: iwwong@hku.hk (W.I. Wong).
[1] Present address: Department of Psychology, University of Hong Kong, Pokfulam, Hong Kong.

http://dx.doi.org/10.1016/j.yhbeh.2015.10.022
0018-506X/© 2015 Elsevier Inc. All rights reserved.

continue to debate the validity and usefulness of 2D:4D for this purpose (Berenbaum etal., 2009; Cohen-Bendahan etal., 2005; McIntyre, 2006; Puts etal., 2004; Wallen, 2009).

2D:4D and prenatal androgen exposure

Arguments for the use of 2D:4D as a proxy for early androgen exposure rest upon several lines of evidence. The first line is direct evidence of its link with early androgen exposure or androgen receptor sensitivity. For instance, a small study found that right hand 2D:4D (2D:4Dr) correlated negatively with the ratio of testosterone to estradiol in amniotic fluid at mid-gestation in a sample of 29 children when sexwas included as a variable in the multiple regression, although no correlation was seen with testosterone alone (Lutchmaya etal., 2004). Another study (Ventura etal., 2013) found that 2D:4D correlated negatively with amniotic testosterone in newborn females, but not in males. Although 2D:4D has been reported to relate negatively to the length of CAG repeats, a measure of the sensitivity of androgen receptors (Manning etal., 2003), a recent meta-analysis did not support this initial finding (Voracek, 2014).

Findings on 2D:4D in non-human animals have been inconsistent. For example, findings on sex differences in animal 2D:4D have not always reported a larger ratio in females, although this is possibly due, in part, to inter-species differences in physiology (for a review, see McIntyre, 2006). More critically, findings from studies manipulating hormones in non-human species also have been inconsistent. For instance, one study found that both reducing androgens and increasing estrogens in utero increased 2D:4D in mice (Zheng and Cohn, 2011). In contrast, however, another study found that female rhesus monkeys exposed to testosterone in utero had feminized, not masculinized, 2D:4Dr (Abbott etal., 2012).

A third line of research relies on individuals with DSDs. For instance, females with increased prenatal androgen exposure due to congenital adrenal hyperplasia (CAH) have been found to have lower 2D:4D than same-sex controls (Brown etal., 2002; Ökten etal., 2002; Rivas etal., 2014), although a study using left-hand radiographs did not report a similar finding (Buck etal., 2003), perhaps because radiographs are less sensitive to sex differences (Hönekopp and Watson, 2010; Wallen, 2009). Also, genetic males with complete androgen insensitivity syndrome (CAIS) have been found to have larger 2D:4D ratios than male controls (Berenbaum etal., 2009), a finding that also offers support for a relationship between 2D:4D and prenatal androgen exposure. However, there is considerable variability and overlap between genetic males with and without CAIS and the effect (right hand $d = .61$; left hand $d = .35$) is smaller than the sex difference in prenatal androgen concentrations as measured in amniotic fluid ($d = 1.40$ or larger) (e.g.,Auyeung etal., 2009; Finegan etal., 1989; Knickmeyer etal., 2005; van de Beek etal., 2009), and in the presumed difference in effective prenatal androgen exposure between genetic males with and without CAIS (Berenbaum etal., 2009).

Twin studies also have been used to try to examine the relationship between 2D:4D and prenatal androgen exposure. One small study found that females with male co-twins have smaller 2D:4D ratios (van Anders etal., 2006), presumably because of exposure to some of their male twins' androgen prenatally (Miller, 1994), but this finding was not replicated in a second, larger, study (Medland etal., 2008).

In sum, studies of DSDs provide some support for a relationship between 2D:4D and prenatal androgen exposure, but they also suggest that the relationship is not strong. Findings from the few experimental studies of non-human animals and opposite-sex twin studies are mixed. Evidence from studies correlating 2D:4D with normal variations in prenatal androgen exposure in humans is also mixed. Thus 2D:4D may be only a weak marker of prenatal androgen exposure, particularly in the typically-developing population (Berenbaum etal., 2009; Hönekopp and Watson, 2010; Knickmeyer etal., 2011; Wallen, 2009).

Early sex difference and stability in 2D:4D

Reports of a sex difference and stability in 2D:4D early in life may be another reason for the popularity of the measure (Cohen-Bendahan etal., 2005). Although a sex difference in 2D:4D has been found in fetuses (Galis etal., 2010; Malas etal., 2006), most studies measure 2D:4D on the surface of the hand after birth, including one study that reported it to be sex-dimorphic and stable from age 2 to 25 years (Manning etal., 1998). Although these findings have been repeatedly cited in support of the use of 2D:4D, they have not been replicated consistently.

One subsequent study found a sex difference in 2D:4Dr in 72 18-month-olds($d = .55$) (left hand not measured) (Saenz and Alexander, 2013), while another study found that 106 newborns showed a sex difference in 2D:4D on the left hand (2D:4Dl) ($d = .57$) but not the right (Ventura etal., 2013). In addition, several studies of children of about age 2 or younger have not found a sex difference in 2D:4D (e.g.,Alexander etal., 2009 on 41 infants; Knickmeyer etal., 2011 on 364 infants; Lutchmaya etal., 2004 on 33 children). Methodological differences have been noted to affect the measurable sex difference in 2D:4D (Auger and Eustache, 2011; Hönekopp and Watson, 2010; Wallen, 2009). In general, measurements made on a computer are more accurate than those made using calipers (Kemper and Schwerdtfeger, 2009). It also has been suggested that the sex difference is larger on the right than on the left hand, and using methods that press the soft tissue, e.g.,by pressing the hand on a scanner, than using direct measures of the fingers or radiographs of bones (Hönekopp and Watson, 2010; Wallen, 2009). The reason for the variability in infant studies is uncertain, as all of these studies measured the digits with calipers on photocopies or digital scans. However, studies with very small samples seem the least likely to report a sex difference.

Also, the conclusion of stability from young childhood reported by Manning etal. (1998) may have obscured instability across a narrower time frame as it relied on a regression analysis spanning the ages 2 to 25 years. In a longitudinal study of radiographed 2D:4Dl, the sex difference in 2D:4D was not significant until age 9 years (McIntyre etal., 2005), and the ratio has been reported to increase with age in both sexes from age one until at least age 17 years (McIntyre etal., 2005; Trivers etal., 2006). A recent study of 0-, 12- and 24-month-old infants also found some temporal correlations in 2D:4D ($r = .35$ to $.53$) in the first 2 years of life, but the ratio decreased in the first year and then increased in the second year (Knickmeyer etal., 2011).

2D:4D–behavior correlations

Studies correlating 2D:4D with behavioral characteristics suggested to be influenced by prenatal androgens have also produced inconsistent findings (for reviews, see Cohen-Bendahan etal., 2005; Puts etal., 2004). Inconsistencies include finding correlations for some sex-typed behaviors but not for others, finding correlations in one sex only, finding correlations for one hand only, finding correlations in some studies but not in others, and finding unpredicted correlations. Studies of adults illustrate these inconsistencies (e.g.,Puts etal., 2004, 2008; Vermeersch etal., 2008). For example, out of 57 correlations with psychological variables hypothesized to relate to 2D:4D or prenatal androgen exposure, only sexual orientation correlated in predicted directions (in both sexes but with 2D:4Dl only) (Puts etal., 2004). In addition, a large meta-analysis found no relationship between 2D:4D and sex-typedspatial abilities (Puts etal., 2008), although findings from a large internet survey found the predicted correlations (Collaer etal., 2007; Peters etal., 2007).

The current study

Thus, prior research suggests that the usefulness of 2D:4D as a measure of prenatal androgen exposure for adults is limited. 2D:4D might

better reflect prenatal androgen exposure in children than in adults, because children have less postnatal experience that is not related to prenatal androgen exposure (Cohen-Bendahan etal., 2005). However, the size and consistency of the sex difference in 2D:4D, and its relation to behavior, in children are poorly understood. We examined whether 2D:4D is sex-dimorphic and stable as early as 2 years of age. We also evaluated correlations between 2D:4D and several sex-typed behaviors in young children. In addition, and unlike previous studies that evaluated 2D:4D–behavior correlations only once or across different samples, we followed the sample up after 6 to 8 months to provide a rare same-sample, same-measure replicability test.

To increase experimental power, and the chances of detecting correlations between 2D:4D and behavior, we used behavioral measures thatshow large sex differences and that have been related to prenatal androgen exposure (Cohen-Bendahan etal., 2005). Sex-typed childhood behavior as assessed by the Preschool Activities Inventory (PSAI) (Golombok and Rust, 1993a, 1993b) and observed toy preferences show large sex differences (e.g.,*d* between 1.5 and 3, Hines, 2010). They also are influenced by early androgen exposure. For instance, girls with CAH show greater preference for boy-typical toys (e.g.,vehicles) and less preference for girl-typical toys (e.g.,dolls) (Berenbaum and Hines, 1992; Nordenström etal., 2002; Pasterski etal., 2005) and obtain more male-typical scores on the PSAI (Hines etal., 2003b) than same-sex controls. In addition, genetic males with insufficient androgen production or incomplete androgen sensitivity show reduced boy-typical and increased girl-typical toy preferences in observations, parent reports, and self-report (Hines etal., 2003a; Jürgensen etal., 2007). Additionally, girls (Ehrhardt etal., 1977), though not boys (Meyer-Bahlburg etal., 1977), whose mothers took anti-androgenic synthetic progestins, have been found to show reduced male-typical play. Girls' PSAI scores have also been found to correlate negatively with testosterone in maternal blood during pregnancy (Hines etal., 2002). Therefore, we measured sex-typed childhood behavior using the PSAI(Golombok and Rust, 1993a, 1993b) and observed play with sex-typed toys (a train and a doll) to evaluate 2D:4D–behavior correlations.

Method

Participants

One hundred and twenty six children from the United Kingdom took part (56 boys, 70 girls). Mean ages in months were: boys, 28.52 (*SD* = 5.79); girls, 29.22 (*SD* = 5.51). Children were tested on two occasions. For the purpose of analysis, we grouped the children into 3 age groups based on their age at the initial test occasion (T1)—20 to 26 months (25 boys, 25 girls) = youngest group, 27 to 33 months (17 boys, 28 girls) = middle group, and 34 to 40 months (14 boys, 17 girls) = oldest group. Of the children, 80% were White, 13% were of mixed race, 1% was Chinese, 2% were Indian, and the rest were of other ethnicities. Ninety-nine children (40 boys, 59 girls) were tested on the second occasion (T2), 6 to 8 months after T1, when the children were aged 26 to 47 months. Mean ages in months were: boys, 35.40 (*SD* = 5.42); girls, 36.10 (*SD* = 5.78). Of these children, 16 boys and 22 girls belonged to the youngest group, 15 boys and 22 girls belonged to the middle group, and 9 boys and 15 girls belonged to the oldest group.

Measures

2D:4D

Because sex differences in 2D:4D may be more detectable with pressure on the soft tissue of the finger tips (Hönekopp and Watson, 2010; Wallen, 2009), each hand was scanned twice, once with a rice bag, which produces pressure, and once with a paper on top for shading purposes. Scans in which the creases could not be seen clearly were excluded. Each length was measured from the midpoint of the lower crease adjacent to the palm to the tip of the finger in units of pixels using a computer program (Gimp). See Table1 for raw digit lengths (sample sizes vary depending on how many children provided useable scans). The 2D:4D ratio was obtained by dividing the length of the second digit by the length of the fourth digit. Fifty scans of fingers were measured independently by a second rater for each time point. Both raters were blind to the sex of the children during measurement. The ratio of 2D:4D correlated similarly highly between raters for both methods (T1: paper = .930, rice bag = .940) (T2: paper = .937, rice bag = .933). At both time-points, 2D:4D measured with rice bag and paper correlated positively (r = .60–.69) and did not interact with the sex difference. Therefore, to increase reliability, a composite 2D:4D for each hand was obtained by averaging across the two conditions whenever possible. Twenty-two children at T1 and 25 children at T2 did not provide usable data for 2D:4Dr. Nine children at T1 and 6 children at T2 did not provide usable data for 2D:4Dl.

Preschool Activities Inventory (PSAI)

The PSAI (Golombok and Rust, 1993a, 1993b) is a 24-item parent report measure assessing frequency of play in regard to a variety of toys, games and activities on a 5-point scale ranging from "1 — *never*" to "5 — *very often*." Each score is calculated with the formula provided in the standardization studies (Golombok and Rust, 1993a, 1993b):

Score = 48.25 + 1.1 × (the sum of "male" items − the sum of "female" items).

Higher scores reflect more male-typical/lessfemale-typical behavior and lower scores reflect more female-typical/lessmale-typical behavior. The measure has been standardized and validated for young children in several countries (Golombok and Rust, 1993a, 1993b) and showed high test-retest reliability in the current study (r = .85). PSAI scores were available for all but one child at T1, and for all children at T2.

Observed toy preferences

Children's play with a boy-typical toy (a train) and a girl-typical toy (a doll) was observed. This procedure was conducted as part of a larger study that also looked at the effect of color on toy preferences. In one condition, the doll was pink and the train was blue. In another condition, the colors of the train and the doll were reversed. For the current study, the effect of color on toy preferences is not of interest. Because toys with sex-typical colors showed larger sex differences and toys are usually in sex-typical colors, analyses for the current study focused on toy preferences in the sex-typical color condition.[2] Children participated with a parent during play due to their young age. Parents were told that they could interact with the child as they normally would. Parents' responses during play were later coded as positive, negative, or neutral and were found to show no relation to children's toy preferences (Wong and Hines, 2015a). Therefore, parental responses will not be discussed further.

The child's behavior was coded as play with the train, play with the doll and not playing at all. The train and the doll were coded separately if they were played with at the same time. The play session lasted for 4 min, although only the first 3 min were coded unless parts of the first 3 min were not codable. For coding purpose, each session was divided into 36, 5-second intervals. To adjust for individual differences in total play time, play with a certain toy was the proportion of intervals playing with that toy out of the total number of intervals playing with any toy. Twenty recordings (40 sessions) were coded independently by a second coder. Inter-rater reliabilities were: blue train (r = .95), pink doll (r = .96). Five children at T1 and 2 children at T2 did not provide data for toy preferences.

[2] When analyzed using the toy preferences combined across the two color conditions, results remained the same. Also, 2D:4D did not correlate with preference for pink or blue as assessed by choices of color cards and gender-neutral toys (see color preferences assessment in Wong and Hines, 2015b) in girls or in boys at either time point.

Table1
Raw digit length in unit of pixels.

		2nd digit				4th digit			
		Rice bag		Paper		Rice bag		Paper	
		Right	Left	Right	Left	Right	Left	Right	Left
T1	Boys	951 (67.21) (39)	948 (68.74) (39)	944 (66.74) (42)	945 (61.23) (40)	1032 (89.94) (39)	1029 (79.71) (39)	1031 (78.95) (42)	1029 (78.53) (40)
	Girls	953 (63.17) (56)	950 (60.24) (55)	954 (53.23) (52)	954 (58.31) (51)	1019 (65.33) (56)	1014 (64.76) (55)	1031 (60.85) (53)	1029 (70.34) (52)
T2	Boys	973 (67.21) (36)	981 (74.56) (34)	984 (66.82) (32)	981 (59.81) (32)	1058 (87.46) (35)	1051 (87.08) (35)	1064 (85.48) (33)	1061 (75.00) (32)
	Girls	995 (66.79) (57)	992 (67.63) (57)	991 (70.21) (54)	991 (72.94) (55)	1058 (74.45) (57)	1052 (71.82) (57)	1058 (76.64) (55)	1053 (72.20) (54)

Note. Data are mean (*SD*) (*n*).

Statistical analyses

The boys and girls did not differ significantly on parent age, parent education or the numbers of male or female siblings; these variables also did not correlate with 2D:4D or the behavioral variables. They were therefore not included as covariates. Missing data were replaced with Expectation Maximization using SPSS Missing Values Analysis with 2500 iterations and a convergence criterion of .0001. This method of replacement is considered one of the best modern model-based methods, superior to traditional methods such as mean substitution (Do and Batzoglou, 2008; Rubin etal., 2007) and is compatible with our analyses (Allison, 2001). Analyses 1 explored the sex difference and stability in 2D:4D across the 6- to 8-month time frame using a 2 (Sex) × 2 (Hand) × 3 (Age group) × 2 (Time) ANOVA. Temporal correlations (*r*) for 2D:4D were calculated to provide further information on stability. Analyses 2 then examined correlations between 2D:4D and the behavioral variables at each time point. To maximize sample size, data for all children were analyzed. However, because ethnicity may relate to the magnitude of the sex difference in 2D:4D (Loehlin etal., 2006), as well as relationships between 2D:4D and behavior (Grimbos etal., 2010), for analyses concerning sex differences in 2D:4D or 2D:4D–behaviorcorrelations, we also examined whether considering ethnicity would change the conclusions. We did not apply Bonferroni adjustments because the primary aim was to evaluate the consistency of correlations between behavior and 2D:4D, so that overly strict Type I error rates would not provide a fair test. Also, Bonferroni adjustments are suitable only for unplanned post-hoc comparisons that address the same hypothesis (Kusuoka and Hoffman, 2002; Rice, 1989).

Results

Analyses 1: sex difference, stability and overlap/variability

Table2 shows means, standard deviations and Ns for the 2 (Sex) × 2 (Hand) × 3 (Age group) × 2 (Time) ANOVA. Table3 shows supplementary descriptive statistics for 2D:4D, including Cohen's *d* for sex differences for each hand at each time point. There were no main effects of Hand or Age group and no interactions. However, there was a main effect of Time, with 2D:4D being larger at T2 than at T1, $F(1, 120) = 23.00$, $p \le .001$, $\eta^2 = .04$. The effect of greatest interest, the main effect of Sex, also was significant, with boys having lower 2D:4D than girls, $F(1, 120) = 12.01$, $p \le .001$, $\eta^2 = .09$ (overall $d = -.61$). Fig. 1 shows substantial within-sex variability and between-sex overlap in 2D:4D at both time points, however.

Temporal correlations (*r*) between T1 and T2 further tested the stability of 2D:4D. The correlations were large in boys for the right hand, $r = .87$, $p \le .001$, and the left hand, $r = .79$, $p \le .001$, as well as in girls for the right hand, $r = .75$, $p \le .001$, and the left hand, $r = .81$, $p \le .001$.

An ANCOVA controlling for ethnicity (White/non-White) produced similar results to the main analyses. There was a main effect of Time, $F(1, 119) = 18.06$, $p < .001$, $\eta^2 = .03$, and a main effect of Sex, $F(1, 119) = 12.30$, $p = .001$, $\eta^2 = .09$ (overall $d = -.67$). Temporal correlations also remained high when ethnicity was controlled (*r*s = .76 to .86, all *p*s < .001).

Analyses 2: behavioral correlations

Analyses 2 evaluated the correlations between 2D:4D and the 3 sex-typed behaviors. We first explored whether these sex-typed behaviors showed the expected sex differences.

Planned *t*-tests showed the expected sex differences for all the behaviors (see Fig.2). Specifically, at T1, boys scored higher than girls on the PSAI,$t(124) = 11.09$, $p < .001$, $d = 1.99$, and showed greater preference than girls for the boy-typical toy, $t(108) = 5.14$, $p < .001$, $d = .81$, unequal variances, and less preference than girls for the girl-typical toy, $t(124) = 5.03$, $p < .001$, $d = -.86$, unequal variances. Similarly, at T2, boys scored higher than girls on the PSAI,$t(99) = 12.28$, $p < .001$, $d = 2.20$, and showed greater preference than girls for the boy-typical toy, $t(110) = 3.91$, $p < .001$, $d = .68$, unequal variances, and less preference than girls for the girl-typical toy, $t(122) = 3.15$, $p = .001$, $d = -.55$, unequal variances.

Correlations between 2D:4D and the 3 sex-typed behaviors are shown in Table4. We hypothesized that 2D:4D would correlate negatively with variables on which boys scored higher (i.e.,PSAI scores and preference for the boy-typical toy) and positively with variables on which girls scored higher (i.e.,preference for the girl-typical toy).

Three out of the 24 correlations with individual behaviors (3 sex-typed behaviors × 2 sexes × 2 time points × 2 hands) were significant and in the expected directions, all of which involved the PSAI at T2. Specifically, PSAI scores at T2 correlated negatively with girls' 2D:4Dr as well as 2D:4Dl, but only with boys' 2D:4Dr and not 2D:4Dl. Contrary to our hypothesis, sex-typed toy preferences did not correlate as

Table2
ANOVA: Mean 2D:4D, SD and N broken down by sex, time, age group and hand.

	2D:4D	*SD*	*N*
Boys	.922	.02	56
Girls	.938	.03	70
Time point 1	.927	.02	126
Time point 2	.933	.02	126
Youngest	.928	.02	50
Middle	.931	.03	45
Oldest	.931	.02	31
Right hand	.928	.03	126
Left hand	.932	.02	126

Table3
Descriptive statistics for 2D:4D and Cohen's d for each hand at each time point.

	Boys ($n = 56$)		Girls ($n = 70$)		
	M	*SD*	*M*	*SD*	*d*
T1					
Right	.920	.029	.933	.031	−.43
Left	.922	.024	.934	.028	−.46
T2					
Right	.921	.026	.939	.028	−.66
Left	.929	.023	.942	.025	−.54

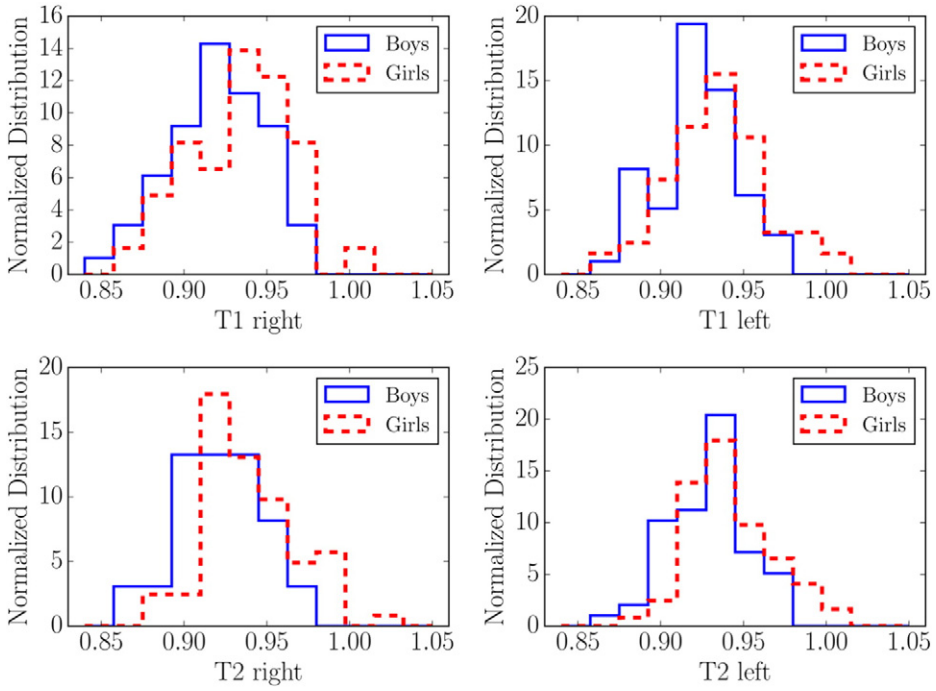

Fig.1. Variability and overlap in boys' and girls' 2D:4D. Overall sex difference $d = .61$, large overlap.

predicted with 2D:4D. Instead, there was an unpredicted significant negative correlation between boys' T2 preference for the girl-typical toy and 2D:4Dr. Results were similar when controlling for ethnicity—PSAI scores at T2 correlated negatively with girls' 2D:4Dr, $r(67) = -.34$, $p = .005$, and 2D:4Dl, $r(67) = -.29$, $p = .014$, but only with boys' 2D:4Dr, $r(53) = -.36$, $p = .007$ and not 2D:4Dl. In addition, boys' T2 preference for the girl-typical toy and 2D:4Dr correlated in the direction opposite to prediction, $r(53) = -.29$, $p = .035$.

We also created a composite measure of sex-typed behavior by summing the standardized scores for the PSAI, preference for the boy-typical toy, and preference for the girl-typical toy (reverse coded so that higher scores are more boy-typical, as is the case for the other two measures). The sex difference in the composite score was very large at T1, $t(122) = 9.24$, $p < .001$, $d = 1.59$, unequal variances, and at T2, $t(110) = 8.32$, $p < .001$, $d = 1.40$, unequal variances, but smaller than the sex difference in PSAI scores (see Fig.2). Its correlation with 2D:4D was significant only for girls' 2D:4Dr at T2 (see Table4).

Because there were 3 significant correlations with the PSAI at T2 and none at T1, and because the PSAI was validated in children aged 2 to 7 years (Golombok and Rust, 1993a, 1993b), we explored whether the inclusion of children aged below 24 months masked any correlations between PSAI scores and 2D:4D at T1. This did not seem to be the case. The correlation between T1 and T2 PSAI scores was high, $r = .88$, $p < .001$. Excluding the 22 children under 24 months of age at T1, the PSAI yielded a sex difference ($d = 2.00$) very close to that including those children ($d = 1.99$).The correlations between T1 PSAI scores and 2D:4D also remained non-significant (though all were negative as predicted).

Finally, we explored the relationship between the magnitude of the sex differences in the behavioral measures and the strength of their correlation with 2D:4D controlling for hand, sex and time. This analysis included 32 correlations (PSAI scores, preference for the boy-typical toy, preference for the girl-typical toy and the behavior composite, ×2 sexes × 2 time points × 2 hands). This correlation between the magnitude of the behavioral sex difference and the relation to 2D:4D was significant and positive, $r(27) = .50$, $p = .006$ (see Fig.3).

Discussion

This study aimed to evaluate 2D:4D as a biomarker of prenatal androgen exposure in young children, by assessing the magnitude of its sex difference, its stability and variability, and its correlations with sex-typed behaviors. We will discuss these findings particularly as

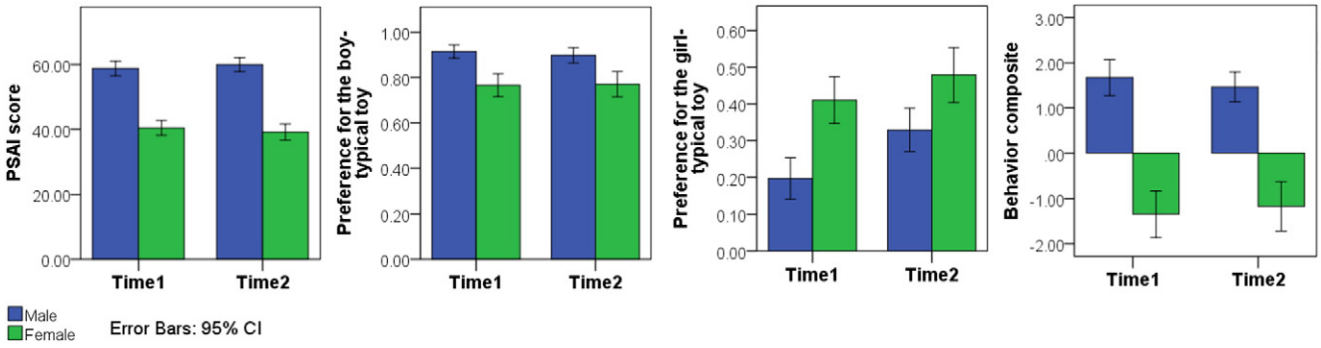

Fig.2. Means and 95% confidence intervals of boys' and girls' PSAI scores, preference for the boy-typical toy, preference for the girl-typical toy, and behavior composite score (higher scores are more boy-typical).

Table4
Correlations between 2D:4D and sex-typed behaviors.

	Time 1				Time 2			
	Boys ($n = 56$)		Girls ($n = 70$)		Boys ($n = 56$)		Girls ($n = 70$)	
	Right	Left	Right	Left	Right	Left	Right	Left
PSAI	$r = -.11$	$r = -.02$	$r = -.05$	$r = -.10$	**$r = -.37^{**}$**	$r = -.19$	**$r = -.34^{**}$**	**$r = -.30^{**}$**
	$p = .428$	$p = .902$	$p = .684$	$p = .398$	$p = .006$	$p = .158$	$p = .004$	$p = .014$
Boy-typical toy	$r = .04$	$r = .16$	$r = .04$	$r = -.02$	$r = .07$	$r = -.11$	$r = -.21$	$r = -.10$
	$p = .766$	$p = .248$	$p = 724$	$p = .852$	$p = .606$	$p = .434$	$p = .078$	$p = .426$
Girl-typical toy	$r = .01$	$r = .05$	$r = .05$	$r = .11$	$r = -.28^{*}$	$r = -.04$	$r = .12$	$r = .06$
	$p = .942$	$p = .710$	$p = .672$	$p = .370$	$p = .034$	$p = .748$	$p = .320$	$p = .604$
Behavior composite	$r = -.04$	$r = .03$	$r = -.02$	$r = -.18$	$r = .04$	$r = -.12$	**$r = -.28^{*}$**	$r = -.18$
	$p = .784$	$p = .841$	$p = .876$	$p = .146$	$p = .775$	$p = .386$	$p = .021$	$p = .147$

Coefficients in bold are significant and in expected directions. Coefficient double underlined is significant and in unexpected direction.
* $p < .05$.
** $p < .005$.

they relate to the usefulness of 2D:4D as a biomarker of prenatal androgen exposure in young children.

Sex difference, stability and variability in 2D:4D

Our results showing a significant sex difference in 2D:4D at mean ages 29 months (20–40 months) and 36 months (26–47 months) replicate Manning etal.'s (1998) finding that 2D:4D is sex-dimorphic from as early as about 2 years of age. The sex difference did not differ by age group, hand or time. The overall effect size for the sex difference (d around .60) was moderate and at the upper end of the range suggested by a meta-analysis of sex differences in 2D:4D ($d = .09$ to .61 in Hönekopp and Watson, 2010). Our finding of high temporal correlations ($r = .75$ to .87) over 6 to 8 months also replicate Manning etal.'s (1998) finding of early temporal stability in 2D:4D, and was similar to those found in older children over a 4-year period (right hand: $r = .78$; left hand: $r = .79$ in Trivers etal., 2006). Both the sex difference and temporal stability appeared to be larger than found in aprior study of 0-, 12-, and 24-month-old infants (non-significant $d = .13$ to .33, $r = .35$ to .53 in Knickmeyer etal., 2011). Recent studies of 18-month-old infants (Saenz and Alexander, 2013) and newborn infants (Ventura etal., 2013) reported a sex difference of similar magnitude, though these latter studies found it only on one hand or measured only one hand.

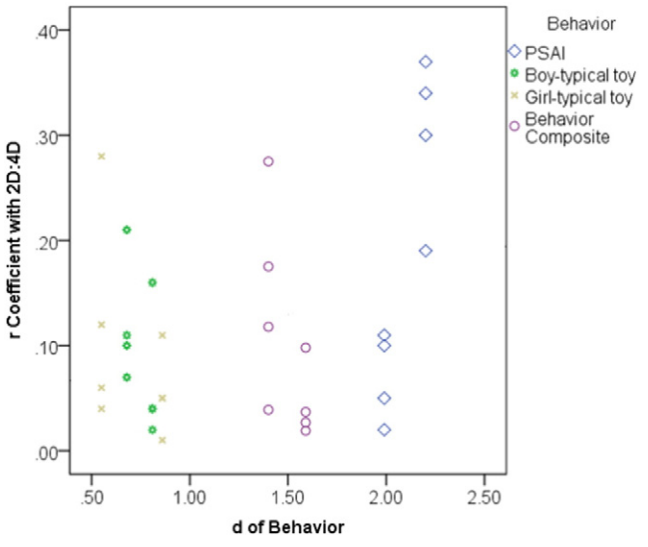

Fig.3. Correlation coefficients tended to increase when the sex differences (in Cohen's d) of the behavioral variables increased, though for each given d, there was variation in the correlation coefficients.

Some studies have found no sex difference in 2D:4D in young children of similar ages. This may have resulted from methodological issues, e.g.,small sample size (e.g.,Lutchmaya etal., 2004), use of Vernier calipers (e.g.,Alexander etal., 2009; Knickmeyer etal., 2011; Lutchmaya etal., 2004), and/or using left hand radiographs of bones (e.g.,McIntyre etal., 2005). In contrast, our study included a moderately large sample size, employed computerized measurements of digitalized scans, which are more precise than non-computerized measurements (Kemper and Schwerdtfeger, 2009), measured both hands, because 2D:4Dr appears to show a larger sex difference than 2D:4Dl (Hönekopp and Watson, 2010), and involved surface measurements, which increases the odds to detect the sex difference, possibly by taking soft tissues into account (Hönekopp and Watson, 2010; Wallen, 2009).

One of the studies that found no consistent sex difference and low stability in infants (Knickmeyer etal., 2011) suggested that studies involving children under the age of two years should include very narrow age ranges. Knickmeyer etal. (2011) also found 2D:4D to first decrease from 0 to 12 months of age, then increase from 12 to 24 months of age. We found a main effect of time that suggests a continued increase in 2D:4D in both boys and girls after that age range. Studies of older children also have reported increased 2D:4D with age and suggest that the increase is due to the relatively faster growth of the second finger compared to the fourth finger (Bloom etal., 2010; McIntyre etal., 2005; Trivers etal., 2006). However, this age-related increase was similar for boys and girls and did not interact with the sex difference. Temporal correlations suggest good stability in 2D:4D beginning at about the age of 2 years. These findings suggest that 2D:4D studies involving children above the age of 2 years can include a wider age range than studies of younger infants. However, it should be noted that the high stability we found was for measurements separated by 6 to 8 months. Future studies might substantiate the early temporal stability in 2D:4D over longer periods of time.

Consistent with others (Berenbaum etal., 2009; Knickmeyer etal., 2011), we also found substantial within-sex variability and between-sex overlap in 2D:4D. The use of 2D:4D as a measure of prenatal androgen exposure rests on the hypothesis that prenatal androgen exposure causes the sex difference in 2D:4D. It has been argued that if 2D:4D is a good measure of fetal testosterone exposure, then the sizes of its sex difference and variability should resemble those in fetal testosterone levels (e.g.,Berenbaum etal., 2009; Hönekopp and Watson, 2010). This does not seem to be the case, however. Studies of amniotic testosterone levels (e.g.,Auyeung etal., 2009; Finegan etal., 1989; Knickmeyer etal., 2005; Lutchmaya etal., 2004; van de Beek etal., 2009) during the hypothesized critical period for human sexual differentiation have found very large sex differences ($d = 1.40$ or larger). Testosterone concentrations in the gonads and in plasma of male and female fetuses show even more marked differences, being about 15 times and 4.5 times as large, respectively, in male compared to female fetuses (Reyes etal., 1973, 1974). In contrast, our results

suggest that 2D:4D in young children shows only moderate-sized sex differences (overall *d* of around .60). This contrast suggests that 2D:4D is a weak reflection of prenatal testosterone exposure (Hönekopp and Watson, 2010; Knickmeyer etal., 2011), limiting its validity for assessing hormone–behavior associations, even in young children.

Behavioral correlates of 2D:4D

All sex-typed behavioral variables showed moderate to very large sex differences at both time points ($d = .55$ to 2.20). Finding large sex differences in the behavioral variables and a moderate sex difference in 2D:4D is important because they enhance sensitivity to sex-related variations. Indeed, our results suggest that correlations between 2D:4D and behavior are stronger if the behaviors show larger sex differences, though for each given size of sex difference, there was some variation in the correlation coefficients.

Out of the 24 correlations between 2D:4D and the three individual behaviors at two time points in two sexes and on two hands, one correlation in the opposite direction to that predicted was found. This correlation involved 2D:4Dr and play with the doll in boys at T2. In addition to this unexpected correlation, 3 significant correlations in expected directions were found, all of which involved the PSAI at T2. The significant predicted effects for the PSAI at T2, but not T1, may have occurred partly because the sex differences in PSAI scores and 2D:4D were larger at T2.

One prior study of children aged 1.9 to 5.6 years found that PSAI scores related negatively to left hand 2D:4D in boys though there was no correlation in girls (Hönekopp and Thierfelder, 2009). However, it is hard to compare the correlations with PSAI scores in this study with those in Hönekopp and Thierfelder's (2009) because they did not use the standardized formula for scoring the PSAI(Golombok and Rust, 1993a, 1993b). Interpretation of their results is also hampered by the lack of a significant sex difference in 2D:4D in their study. The fact that both studies found negative correlations between 2D:4D and PSAI scores could be interpreted to support the use of 2D:4D for assessing behavioral effects of early androgen exposure. However, it is unclear why the only correlation found in their study (i.e.,with left hand 2D:4D in boys) was the only correlation not found in the current study.

In sum, the significant correlations with PSAI scores partially replicate a previous finding (Hönekopp and Thierfelder, 2009), but the exact finding reported in that study was not replicated in our data set. In addition, as in studies of adults (e.g.,Puts etal., 2004, 2008; Vermeersch etal., 2008), the overall pattern of correlations in our study exhibits a number of inconsistencies, including 1)not showing similar patterns of significance in both sexes, 2)not showing significance for all behaviors previously found to relate to prenatal androgen exposure, 3)not showing significance for both hands even though 2D:4D did not differ significantly by hand, and 4)for the PSAI scores, not showing significance at both time points even though the samples and measures used were identical. Therefore, although children are likely to have experienced fewer postnatal influences on behavior than adults, 2D:4D does not appear to be a better marker of prenatal androgen exposure in young children than in older samples.

Strengths and limitations

We tested the same sample using the same measures twice, separated by 6 to 8 months. This provides a rare replicability test that may be more desirable than different-sample tests, because many sample and measurement characteristics are matched across tests. However, although replication on the same sample can be useful, same-samplenon-replicability may be caused by developmental differences, or by the effects of experience with the tasks. Different-sample replications could further substantiate our findings.

Our behavioral measures all showed moderate to very large sex differences and our 2D:4D measures showed moderate sex differences. Nevertheless, we found few significant correlations between 2D:4D and sex-typed behaviors. More correlations might have been seen if larger samples were studied. It should be noted, however, that most of the non-significant correlation coefficients were close to zero, and many were in unpredicted directions. Similarly, more correlations might have been seen if the behavioral variables showed even larger sex differences. PSAI scores showed the largest sex difference of all the behavioral variables and they showed some predicted correlations with 2D:4D. In addition, the size of the correlations between 2D:4D and behavioral variables correlated positively with the size of the sex differences in the behavioral variables. Large sex differences in behavioral measures may increase the power of studies trying to detect correlations between 2D:4D and behavior. Although the toy preference variables in the current study showed large sex differences, some studies have found even larger sex differences in toy preferences (e.g.,Pasterski etal., 2005), so future studies might usefully assess toy preferences using toys that elicit even larger sex differences. The sex differences in toy preferences also grow larger across childhood and so studies of older children might also provide more powerful tests.

Interpreting 2D:4D–behavior correlations and suggestions for future research

Explanations of inconsistencies in the findings of correlations between 2D:4D and behavior are largely speculative. Explanations include the suggestions that 2D:4D on a certain hand, in relation to a certain behavior or for a certain sex may be particularly sensitive to early androgenic influence. For example, findings that 2D:4D shows behavioral correlations only for 2D:4Dr have been explained by suggesting a lateralized effect of androgens, such that androgens affect the right side of the body more than the left (e.g.,Manning etal., 1998). There is no direct evidence that androgens affect 2D:4Dr more than 2D:4Dl, however. Similarly, findings that 2D:4D correlations are found only forthe left hand have led to suggestions that 2D:4Dr and 2D:4Dl are influenced by androgens at different times prenatally, and so relate to different behaviors (e.g.,Hönekopp and Thierfelder, 2009). However, a meta-analytic investigation of the behavioral associations for 2D:4Dr versus 2D:4Dl did not find 2D:4D on the two hands to differ consistently in their ability to predict behavior (Hönekopp and Schuster, 2010 on 2D:4D-athletic prowess correlations). Similarly, it has been suggested that the behaviors that correlate with 2D:4D have similar critical periods for androgen influence as 2D:4D and those that do not correlate with 2D:4D have different critical periods to that for 2D:4D (e.g.,Puts etal., 2004). While this explanation is possible, not enough is known about the critical periods for androgenic influence on specific human behaviors to evaluate its validity. A final explanation for inconsistencies in results relates to inconsistencies in methodologies for assessing 2D:4D or for assessing behavior. Whereas such methodological inconsistencies may be reasonable explanations of different findings in different studies, it is harder to reconcile inconsistent findings in the current study, because we used the same procedures, measures and samples.

More research into the details of prenatal androgen mechanisms in general and of 2D:4D in particular, such as the critical period of influence for various behaviors, and possible different mechanisms and timing of influence for the left versus right hand and for males versus females, would help in interpreting correlations involving 2D:4D. Ameta-analysis on the sex difference in 2D:4D (Hönekopp and Watson, 2010) has provided useful information on the size of the sex differences. Additional meta-analyses on 2D:4D–behavior associations similar to existing ones on sexual orientation (Grimbos etal., 2010), spatial abilities (Puts etal., 2008) and athletic prowess (Hönekopp and Schuster, 2010) could also be helpful.

It also is possible that other approaches to measuring the early hormone environment in typically-developing individuals are preferable to 2D:4D. For instance, AGD shows a larger sex difference than 2D:4D and there is consistent evidence that it relates to early androgen exposure in humans and in rodents (Callegari etal., 1987; Welsh etal., 2008; Yeh

etal., 2002). In addition, AGD at birth has been found to relate as predicted to PSAI scores for boys at age 4 years, although not for girls (Pasterski etal., 2015). Although measures of AGD may be difficult toobtain because it is not routinely measured at birth and may be considered sensitive in older samples, it may be necessary to use more challenging measures to obtain reliable results.

Acknowledgment

This research was funded by the Gates Cambridge Trust and St. John's College, University of Cambridge.

References

Abbott, A.D., Colman, R.J., Tiefenthaler, R., Dumesic, D.A., Abbott, D.H., 2012. Early-to-mid gestation fetal testosterone increases right hand 2D:4D finger length ratio in polycystic ovary syndrome-like monkeys. PLoS One 7. http://dx.doi.org/10.1371/journal.pone.0042372.

Alexander, G.M., Wilcox, T., Farmer, M.E., 2009. Hormone–behavior associations in early infancy. Horm. Behav. 56, 498–502. http://dx.doi.org/10.1016/j.yhbeh.2009.08.003.

Allison, P.D., 2001. Missing Data. Sage, Thousand Oaks, CA.

Auger, J., Eustache, F., 2011. Second to fourth digit ratios, male genital development and reproductive health: a clinical study among fertile men and testis cancer patients. Int. J. Androl. 34. http://dx.doi.org/10.1111/j.1365-2605.2010.01124.x.

Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K., Hackett, G., Hines, M., 2009. Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. Psychol. Sci. 20, 144–148. http://dx.doi.org/10.1111/j.1467-9280.2009.02279.x.

Bailey, A.A., Hurd, P.L., 2005. Finger length ratio (2D :4D) correlates with physical aggression in men but not in women. Biol. Psychol. 68, 215–222. http://dx.doi.org/10.1016/j.biopsycho.2004.05.001.

Berenbaum, S.A., Bryk, K.K., Nowak, N., Quigley, C.A., Moffat, S., 2009. Fingers as a marker of prenatal androgen exposure. Endocrinology 150, 5119–5124. http://dx.doi.org/10.1210/en.2009-0774.

Berenbaum, S.A., Hines, M., 1992. Early androgens are related to childhood sex-typed toy preferences. Psychol. Sci. doi: http://dx.doi.org/10.1111/j.1467-9280.1992.tb00028.x.

Bloom, M.S., Houston, A.S., Mills, J.L., Molloy, C.A., Hediger, M.L., 2010. Finger bone immaturity and 2D:4D ratio measurement error in the assessment of the hyperandrogenic hypothesis for the etiology of autism spectrum disorders. Physiol. Behav. 100, 221–224. http://dx.doi.org/10.1016/j.physbeh.2010.01.005.

Brosnan, M.J., 2008. Digit ratio as an indicator of numeracy relative to literacy in 7-year-old British schoolchildren. Br. J. Psychol. 99, 75–85. http://dx.doi.org/10.1348/000712607X197406.

Brown, W.M., Hines, M., Fane, B.A., Breedlove, S.M., 2002. Masculinized finger length patterns in human males and females with congenital adrenal hyperplasia. Horm. Behav. 42, 380–386. http://dx.doi.org/10.1006/hbeh.2002.1830.

Buck, J.J., Williams, R.M., Hughes, I.A., Acerini, C.L., 2003. In-utero androgen exposure and 2nd to 4th digit length ratio — comparisons between healthy controls and females with classical congenital adrenal hyperplasia. Hum. Reprod. 18, 976–979. http://dx.doi.org/10.1093/humrep/deg198.

Callegari, C., Everett, S., Ross, M., Brasel, J.A., 1987. Anogenital ratio: measure of fetal virilization in premature and full-term newborn infants. J. Pediatr. 111, 240–243.

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. Neurosci. Biobehav. Rev. 29, 353–384. http://dx.doi.org/10.1016/j.neubiorev.2004.11.004.

Collaer, M.L., Reimers, S., Manning, J.T., 2007. Visuospatial performance on an internet line judgment task and potential hormonal markers: sex, sexual orientation, and 2D:4D. Arch. Sex. Behav. 36, 177–192. http://dx.doi.org/10.1007/s10508-006-9152-1.

Dean, A., Sharpe, R.M., 2013. Anogenital distance or digit length ratio as measures of fetal androgen exposure: relationship to male reproductive development and its disorders. J. Clin. Endocrinol. Metab. 98, 2230–2238. http://dx.doi.org/10.1210/jc.2012-4057.

Do, C.B., Batzoglou, S., 2008. What is the expectation maximization algorithm? Nat. Biotechnol. 26, 897–899. http://dx.doi.org/10.1038/nbt1406.

Ehrhardt, A.A., Grisanti, G.C., Meyer-Bahlburg, H.F.L., 1977. Prenatal exposure to medroxyprogesterone acetate (MPA) in girls. Psychoneuroendocrinology 2, 391–398. http://dx.doi.org/10.1016/0306-4530(77)90010-5.

Finegan, J.A., 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. http://dx.doi.org/10.1080/00221325.1989.9914580.

Galis, F., Ten Broek, C.M.A., Van Dongen, S., Wijnaendts, L.C.D., 2010. Sexual dimorphism in the prenatal digit ratio (2D:4D). Arch. Sex. Behav. 39, 57–62. http://dx.doi.org/10.1007/s10508-009-9485-7.

Golombok, S., Rust, J., 1993a. The measurement of gender role behavior in preschoolbehavior. J. Child Psychol. Psychiatry 34, 805–811.

Golombok, S., Rust, J., 1993b. The pre-school activities inventory: a standardized assessment of gender role in children. Psychol. Assess. 5, 131–136. http://dx.doi.org/10.1037/1040-3590.5.2.131.

Grimbos, T., Dawood, K., Burriss, R.P., Zucker, K.J., Puts, D.A., 2010. Sexual orientation and the second to fourth finger length ratio: a meta-analysis in men and women. Behav. Neurosci. 124, 278–287. http://dx.doi.org/10.1037/a0018764.

Hines, M., 2010. Sex-related variation in human behavior and the brain. Trends Cogn. Sci. 14, 448–456. http://dx.doi.org/10.1016/j.tics.2010.07.005.

Hines, M., Ahmed, S.F., Hughes, I.A., 2003a. Psychological outcomes and gender-related development in complete androgen insensitivity syndrome. Arch. Sex. Behav. 32, 93–101. http://dx.doi.org/10.1023/A:1022492106974.

Hines, M., Fane, B.A., Pasterski, V.L., Mathews, G.A., Conway, G.S., Brook, C., 2003b. Spatial abilities following prenatal androgen abnormality: targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. Psychoneuroendocrinology 28, 1010–1026. http://dx.doi.org/10.1016/S0306-4530(02)00121-X.

Hines, M., Golombok, S., Rust, J., Johnston, K.J., Golding, J., 2002. Testosterone during pregnancy and gender role behavior of preschool children: a longitudinal, population study. Child Dev. 73, 1678–87. doi: http://dx.doi.org/http://dx.doi.org/10.1111/1467-8624.00498.

Hönekopp, J., Schuster, M., 2010. A meta-analysis on 2D:4D and athletic prowess: substantial relationships but neither hand out-predicts the other. Personal. Individ. Differ. 48, 4–10. http://dx.doi.org/10.1016/j.paid.2009.08.009.

Hönekopp, J., Thierfelder, C., 2009. Relationships between digit ratio (2D:4D) and sex-typed play behavior in pre-school children. Personal. Individ. Differ. 47, 706–710. http://dx.doi.org/10.1016/j.paid.2009.06.007.

Hönekopp, J., Watson, S., 2010. Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand. Am. J. Hum. Biol. 22, 619–630. http://dx.doi.org/10.1002/ajhb.21054.

Jürgensen, M., Hiort, O., Holterhus, P.-M., Thyen, U., 2007. Gender role behavior in children with XY karyotype and disorders of sex development. Horm. Behav. 51, 443–453. http://dx.doi.org/10.1016/j.yhbeh.2007.01.001.

Kemper, C.J., Schwerdtfeger, A., 2009. Comparing indirect methods of digit ratio (2D:4D) measurement. Am. J. Hum. Biol. 21, 188–191. http://dx.doi.org/10.1002/ajhb.20843.

Knickmeyer, R.C., Wheelwright, S., Taylor, K., Raggatt, P., Hackett, G., Baron-Cohen, S., 2005. Gender-typed play and amniotic testosterone. Dev. Psychol. 41, 517–528. http://dx.doi.org/10.1037/0012-1649.41.3.517.

Knickmeyer, R.C., Woolson, S., Hamer, R.M., Konneker, T., Gilmore, J.H., 2011. 2D:4D ratios in the first 2 years of life: stability and relation to testosterone exposure and sensitivity. Horm. Behav. 60, 256–263. http://dx.doi.org/10.1016/j.yhbeh.2011.05.009.

Kusuoka, H., Hoffman, J.I.E., 2002. Advice on statistical analysis for circulation research. Circ. Res. 91, 662–671. http://dx.doi.org/10.1161/01.RES.0000037427.73184.C1.

Loehlin, J.C., McFadden, D., Medland, S.E., Martin, N.G., 2006. Population differences in finger-length ratios: ethnicity or latitude? Arch. Sex. Behav. 35, 739–742. http://dx.doi.org/10.1007/s10508-006-9039-1.

Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., Manning, J.T., 2004. 2nd to 4th digit ratios, fetal testosterone and estradiol. Early Hum. Dev. 77, 23–28. http://dx.doi.org/10.1016/j.earlhumdev.2003.12.002.

Malas, M.A., Dogan, S., Hilal Evcil, E., Desdicioglu, K., 2006. Fetal development of the hand, digits and digit ratio (2D :4D). Early Hum. Dev. 82, 469–475. http://dx.doi.org/10.1016/j.earlhumdev.2005.12.002.

Manning, J.T., Baron-Cohen, S., Wheelwright, S., Sanders, G., 2001. The 2nd to 4th digit ratio and autism. Dev. Med. Child Neurol. 43, 160–164. http://dx.doi.org/10.1097/00004703-200110000-00019.

Manning, J.T., Bundred, P.E., Newton, D.J., Flanagan, B.F., 2003. The second to fourth digit ratio and variation in the androgen receptor gene. Evol. Hum. Behav. 24, 399–405. http://dx.doi.org/10.1016/S1090-5138(03)00052-7.

Manning, J.T., Scutt, D., Wilson, J., Lewis-Jones, D.I., 1998. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. Hum. Reprod. 13, 3000–3004. http://dx.doi.org/10.1093/humrep/13.11.3000.

McIntyre, M.H., 2006. The use of digit ratios as markers for perinatal androgen action. Reprod. Biol. Endocrinol. 4, 10. http://dx.doi.org/10.1186/1477-7827-4-10.

Mcintyre, M.H., Ellison, P.T., Lieberman, D.E., Demerath, E., Towne, B., 2005. The development of sex differences in digital formula from infancy in the Fels Longitudinal Study. Proc. R. Soc. B 272, 1473–1479. http://dx.doi.org/10.1098/rspb.2005.3100.

Medland, S.E., Loehlin, J.C., Martin, N.G., 2008. No effects of prenatal hormone transfer on digit ratio in a large sample of same- and opposite-sex dizygotic twins. Personal. Individ. Differ. 44, 1225–1234. http://dx.doi.org/10.1016/j.paid.2007.11.017.

Meyer-Bahlburg, H.F.L., Dolezal, C., Baker, S.W., New, M.I., 2008. Sexual orientation in women with classical or non-classical congenital adrenal hyperplasia as a function of degree of prenatal androgen excess. Arch. Sex. Behav. 37, 85–99. http://dx.doi.org/10.1007/s10508-007-9265-1.

Meyer-Bahlburg, H.F.L., Grisanti, G.C., Ehrhardt, A.A., 1977. Prenatal effects of sex hormones on human male behavior: medroxyprogesterone acetate (MPA). Psychoneuroendocrinology 2, 383–390. http://dx.doi.org/10.1016/0306-4530(77)90009-9.

Miller, E.M., 1994. Prenatal sex hormone transfer: a reason to study opposite-sex twins. Personal. Individ. Differ. 17, 511–529. http://dx.doi.org/10.1016/0191-8869(94)90088-4.

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. J. Clin. Endocrinol. Metab. 87, 5119–5124. http://dx.doi.org/10.1210/jc.2001-011531.

Ökten, A., Kalyoncu, M., Yariş, N., 2002. The ratio of second- and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Early Hum. Dev. 70, 47–54. http://dx.doi.org/10.1016/S0378-3782(02)00073-7.

Pasterski, V., Acerini, C.L., Dunger, D.B., Ong, K.K., Hughes, .I.A, Thankamony, A., Hines, M., 2015. Postnatal penile growth concurrent with mini-puberty predicts later sex-typed play behavior: evidence for neurobehavioral effects of the postnatal androgen surge in typically developing boys. Horm. Behav. 69, 98–105. http://dx.doi.org/10.1016/j.yhbeh.2015.01.002.

Pasterski, V., Geffner, M.E., Hindmarsh, P., Brook, C., Brain, 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 Dev. 76, 264–278. http://dx.doi.org/10.1111/j.1467-8624.2005.00843.x.

Pasterski, V., Hindmarsh, P., Geffner, M., Brook, C., Brain, C., Hines, M., 2007. Increased aggression and activity level in 3- to 11-year-old girls with congenital adrenal hyperplasia (CAH). Horm. Behav. 52, 368–374. http://dx.doi.org/10.1016/j.yhbeh.2007.05.015.

Peters, M., Manning, J.T., Reimers, S., 2007. The effects of sex, sexual orientation, and digit ratio (2D:4D) on mental rotation performance. Arch. Sex. Behav. 36, 251–260. http://dx.doi.org/10.1007/s10508-006-9166-8.

Puts, D.A., McDaniel, M.A., Jordan, C.L., Breedlove, S.M., 2008. Spatial ability and prenatal androgens: meta-analyses of congenital adrenal hyperplasia and digit ratio (2D:4D) studies. Arch. Sex. Behav. 37, 100–111. http://dx.doi.org/10.1007/s10508-007-9271-3.

Puts, D.A., Gaulin, S.J.C., Sporter, R.J., McBurney, D.H., 2004. Sex hormones and finger length — what does 2D:4D indicate? Evol. Hum. Behav. 25, 182–199. http://dx.doi.org/10.1016/j.evolhumbehav.2004.03.005.

Reyes, F.I., Boroditsky, R.S., Winter, J.S.D., Faiman, C., 1974. Studies on human sexual development. II. Fetal and maternal serum gonadotropin and sex steroid concentrations. J. Clin. Endocrinol. Metab. 38, 612–617. http://dx.doi.org/10.1210/jcem-38-4-612.

Reyes, F.I., Winter, J.S.D., Faiman, C., 1973. Studies on human sexual development. I. Fetal gonadal and adrenal sex steroids. J. Clin. Endocrinol. Metab. 37, 74–78.

Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43, 223–225. http://dx.doi.org/10.2307/2409177.

Rivas, M.P., Moreira, L.M.A., Santo, L.D.E., Marques, A.C.S.S., El-Hani, C.N., Toralles, M.B.P., 2014. New studies of second and fourth digit ratio as a morphogenetic trait in subjects with congenital adrenal hyperplasia. Am. J. Hum. Biol. 26, 559–561. http://dx.doi.org/10.1002/ajhb.22545.

Rubin, L.H., Witkiewitz, K., Andre, J.S., Reilly, S., 2007. Methods for handling missing data in the behavioral neurosciences: don't throw the baby rat out with the bath water. J. Undergrad. Neurosci. Educ. 5, A71–A77.

Saenz, J., Alexander, G.M., 2013. Digit ratios (2D:4D), postnatal testosterone and eye contact in toddlers. Biol. Psychol. 94, 106–108. http://dx.doi.org/10.1016/j.biopsycho.2013.05.010.

Salazar-Martinez, E., Romano-Riquer, P., Yanez-Marquez, E., Longnecker, M.P., Hernandez-Avila, M., 2004. Anogenital distance in human male and female newborns: a descriptive, cross-sectional study. Environ. Heal. 3, 8. http://dx.doi.org/10.1186/1476-069X-3-8.

Trivers, R., Manning, J., Jacobson, A., 2006. A longitudinal study of digit ratio (2D:4D) and other finger ratios in Jamaican children. Horm. Behav. 49, 150–156. http://dx.doi.org/10.1016/j.yhbeh.2005.05.023.

van Anders, S.M., Vernon, P.A., Wilbur, C.J., 2006. Finger-length ratios show evidence of prenatal hormone-transfer between opposite-sex twins. Horm. Behav. 49, 315–319. http://dx.doi.org/10.1016/j.yhbeh.2005.08.003.

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. Arch. Sex. Behav. 38, 6–15. http://dx.doi.org/10.1007/s10508-007-9291-z.

Ventura, T., Gomes, M.C., Pita, A., Neto, M.T., Taylor, A., 2013. Digit ratio (2D:4D) in newborns: influences of prenatal testosterone and maternal environment. Early Hum. Dev. 89, 107–112. http://dx.doi.org/10.1016/j.earlhumdev.2012.08.009.

Vermeersch, H., T'Sjoen, G., Kaufman, J.M., Vincke, J., 2008. 2d:4d, sex steroid hormones and human psychological sex differences. Horm. Behav. 54, 340–346. http://dx.doi.org/10.1016/j.yhbeh.2008.02.017.

Voracek, M., 2014. No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D: 4D): a comprehensive meta-analysis and critical evaluation of research. Evol. Hum. Behav. 35, 430–437. http://dx.doi.org/10.1016/j.evolhumbehav.2014.05.009.

Voracek, M., Loibl, L.M., 2009. Scientometric analysis and bibliography of digit ratio (2D: 4D) research, 1998–2008. Psychol. Rep. 104, 922–956. http://dx.doi.org/10.2466/pr0.104.3.922-956.

Wallen, K., 2009. Does finger fat produce sex differences in second to fourth digit ratios? Endocrinology 150, 4819–4822. http://dx.doi.org/10.1210/en.2009-0986.

Welsh, M., Saunders, P.T.K., Fisken, M., Scott, H.M., Hutchison, G.R., Smith, L.B., Sharpe, R.M., 2008. Identification in rats of a programming window for reproductive tract masculinization, disruption of which leads to hypospadias and cryptorchidism. J. Clin. Invest. 118, 1479–1490. http://dx.doi.org/10.1172/JCI34241.

Wong, W.I., Hines, M., 2015a. Effects of gender color-coding on toddlers'gender-typical toy play. Arch. Sex. Behav. 44, 1233–1242. http://dx.doi.org/10.1007/s10508-014-0400-5.

Wong, W.I., Hines, M., 2015b. Preferences for pink and blue: the development of color preferences as a distinct gender-typed behavior in toddlers. Arch. Sex. Behav. 44, 1243–1254. http://dx.doi.org/10.1007/s10508-015-0489-1.

Yeh, S., Tsai, M.-Y., Xu, Q., Mu, X.-M., Lardy, H., Huang, K.-E., Lin, H., Yeh, S.-D., Altuwaijri, S., Zhou, X., Xing, L., Boyce, B.F., Hung, M.-C., Zhang, S., Gan, L., Chang, C., 2002. Generation and characterization of androgen receptor knockout (ARKO) mice: an invivo model for the study of androgen functions in selective tissues. Proc. Natl. Acad. Sci. 99, 13498–13503. http://dx.doi.org/10.1073/pnas.212474399.

Zheng, Z., Cohn, M.J., 2011. From the cover: developmental basis of sexually dimorphic digit ratios. Proc. Natl. Acad. Sci. 108, 16289–16294. http://dx.doi.org/10.1073/pnas.1108312108.

Zucker, K.J., Bradley, S.J., Oliver, G., Blake, J., Fleming, S., Hood, J., 1996. Psychosexual development of women with congenital adrenal hyperplasia. Horm. Behav. 30, 300–318. http://dx.doi.org/10.1006/hbeh.1996.0038.