The influence of testosterone and cortisol on early infant development and the mother-infant-interaction

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Luisa Ernsten

aus Bocholt

Düsseldorf, Mai 2024

aus dem Institut für Experimentelle Psychologie

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Berichterstatter*innen:

1. PD Dr. Nora K. Schaal
2. Prof. Dr. Susanne Becker
3. Prof. Dr. Norbert Zmyj

Tag der mündlichen Prüfung: 27.11.2024

Table of contents

List of abbreviations

2D	Length of the second digit.
3D	Length of the third digit.
4D	Length of the fourth digit.
5D	Length of the fifth digit.
2D:4D	Ratio of the length of the second to the length of the fourth finger
ACTH	Adrenocorticotropic hormone
CAIS	Complete androgen insensitivity syndrome
CAH	Congenital adrenal hyperplasia
CRH	Corticotropin releasing hormone
dpi	Dots per inch
$D_{[r-l]}$	Differential asymmetry; Right 2D:4D subtracted by left 2D:4D
E	Estradiol
e.g.	exempli gratia, engl. for example
FFSF	Face-to-face still-face paradigm
HCC	Hair cortisol concentration
HPA axis	Hypothalamus pituitary adrenal axis
i.e.	id est, engl. that is
rel2	Ratio of the second digit to the cumulated length of all four digits.
rel3	Ratio of the third digit to the cumulated length of all four digits.
rel4	Ratio of the fourth digit to the cumulated length of all four digits.
rel5	Ratio of the fifth digit to the cumulated length of all four digits.
SCC	Salivary cortisol concentration
T	Testosterone
T/E	Ratio of testosterone to estradiol concentration

Zusammenfassung

Hormone spielen als molekulare Botenstoffe eine zentrale Rolle in der Regulation verschiedenster Körperprozesse. Im Rahmen der vorliegenden Arbeit wird die Relevanz zweier ausgewählter Hormone, Testosteron und Cortisol, näher beleuchtet. Testosteron ist bereits pränatal an der Entwicklung von primären Geschlechtsmerkmalen beteiligt und prädisponiert die Entwicklung sekundärer Geschlechtsmerkmale. Im Menschen ist es jedoch nur schwer möglich Testosteron pränatal zu messen, und die Forschung hat stellvertretende Maße, wie das Verhältnis vom zweiten zum vierten Finger (2D:4D), diskutiert. Das 2D:4D zeigt jedoch Schwächen, vor allem hinsichtlich eines direkteren Zusammenhangs zu Testosteron sowie der Anwendbarkeit in Kindern. Cortisol wird mit der menschlichen Stressreaktion in Verbindung gebracht, welche durch die Hypothalamus-Hypophysen-Nebennieren (HPA)-Achse reguliert wird und kann dabei verschiedenste kognitive Prozesse beeinflussen. Im Kontext der frühen Mutter-Kind-Interaktion kann Cortisol potentiell relevante kognitive Ressourcen der Mutter negativ beeinflussen. Es fehlen jedoch standardisierte, experimentelle Studien, die den Zusammenhang zwischen mütterlichem Cortisol und der Qualität der Mutter-Kind-Interaktion untersuchen.

Im ersten Teil der Arbeit wird das Ziel verfolgt die Anwendbarkeit vom 2D:4D in kindlichen Stichproben näher zu untersuchen. Hierzu werden drei Studien vorgestellt, die sowohl die grundlegende Annahme eines Zusammenhangs zwischen dem 2D:4D zu Testosteron-Leveln, als auch den Geschlechterunterschied, der in erwachsenen Stichproben relativ robust scheint sowie letztlich die Anwendbarkeit des 2D:4D als Maß für Testosteron im Zusammenhang mit geschlechtstypischen Verhaltensweisen von Kindern untersuchen.

Studie 1 untersuchte den Zusammenhang zwischen Testosteron gemessen in der Amnionflüssigkeit und dem kindlichen 2D:4D zu mehreren postnatalen Messzeitpunkten vom Säuglings- bis in das Vorschulalter. Aufgrund zu niedriger Testosteronwerte in der Amnionflüssigkeit von weiblichen Kindern konnte der Zusammenhang zum 2D:4D nur in der männlichen Stichprobe untersucht werden, wobei sich zu keinem der Messzeitpunkte ein signifikanter Zusammenhang zeigte. Jedoch

zeigte sich zu jedem der Messzeitpunkte, dass weibliche Kinder ein größeres 2D:4D im Vergleich zu männlichen Kindern aufwiesen. Weiterhin zeigte sich das Alter der Kinder als wichtiger Einflussfaktor, sodass das 2D:4D mit zunehmendem Alter größer wurde. Die Ergebnisse der Studie deuten auf mögliche Schwächen des 2D:4D als Marker für pränatales Testosteron hin und verdeutlichen die Notwendigkeit weiterer Forschung.

Studie 2 untersuchte Geschlechterunterschiede in verschiedenen Hand- und Fingerlängenverhältnissen von Kindern sowie die Reliabilität von Handscans zur Messung diese Verhältnisse. Männliche Kinder hatten insgesamt größere Hände und längere Finger. Das 2D:4D zeigte nur kleine oder nicht signifikante Geschlechterunterschiede, wohingegen Fingerlängenverhältnisse mit dem fünften Finger die größten Unterschiede aufzeigen konnten. Handscans erwiesen sich als reliable Quelle zur Messung von Hand- und Fingerlängenverhältnissen, jedoch ließen die Ergebnisse ableiten, dass Messfehler durch weniger Compliance und mögliche Verzerrungen in den Scans durch einen höheren Körperfettanteil in jungen Kindern im Vergleich zu Erwachsenen auftreten können. Insgesamt unterstützt die Studie Geschlechterunterschiede in den Händen und Fingern von Kindern, das 2D:4D war dabei jedoch das schwächste Maß, was seine Anwendbarkeit vor allem in jüngeren Kohorten möglicherweise erschwert.

In Studie 3 wurde der Zusammenhang vom 2D:4D und geschlechtstypischem Spielverhalten von Kindern untersucht. Weiterhin wurde der Einfluss von Geschwisterkindern zusätzlich berücksichtigt. Es zeigte sich kein signifikanter Unterschied im 2D:4D von männlichen und weiblichen Kindern. Der Zusammenhang zwischen dem 2D:4D und geschlechtstypischem Spielverhalten wurde zudem nur in männlichen Einzelkindern signifikant. In der Gesamtstichprobe zeigten sich dagegen signifikante Einflüsse durch ältere und jüngere Schwestern sowie ältere Brüder, jedoch ebenfalls ausschließlich in männlichen Kindern. Das geschlechtstypische Spielverhalten von weiblichen Kindern hing weder mit dem 2D:4D noch mit dem Einfluss von Geschwisterkindern zusammen. Die Ergebnisse der Studie deuten auf eine mögliche Anwendbarkeit des 2D:4D als Marker für pränatales Testosteron hin, allerdings nur

für Jungen ohne Geschwisterkinder. Zudem weisen die Ergebnisse daraufhin, dass Geschwisterkinder eine wichtige Rolle für das geschlechtstypische Spielverhalten von Kindern spielen.

Der zweite Teil der Arbeit untersuchte den Zusammenhang von mütterlichem Cortisol und der Mutter-Kind-Interaktion in einer standardisierten Verhaltensbeobachtung, dem Still-Face-Paradigma. Im Rahmen von Studie 4 wurde mütterliches Cortisol im Speichel sowie im Haar der Mutter gemessen, welche die situationale HPA-Achsen-Aktivität sowie eine längerfristige Aktivität der HPA-Achse indizieren sollten. Es zeigte sich ein Zusammenhang des Speichelcortisols mit weniger kindlichem positiven Affekt sowie dyadischer Interaktion und Blickkontakt, jedoch mit mehr kindlichem Protestverhalten. Mütterliches Haarcortisol schien nicht mit dem Interaktionsverhalten von Mutter und Kind zusammenzuhängen. Die Ergebnisse deuten darauf hin, dass akute nicht jedoch längerfristige HPA-Achsen-Aktivität die Mutter-Kind-Interaktion in negativer Art und Weise beeinflussen könnte.

Die Ergebnisse der Dissertation im ersten Teil deuten auf relevante Schwächen des 2D:4D in der Anwendung bei Kindern hin. Dabei ist vor allem der fehlende Zusammenhang zum pränatalen Testosteron in Studie 1 sowie der fehlende Geschlechterunterschied im 2D:4D in den Studien 2 und 3 besonders hervorzuheben. Jedoch lieferten alle Studien ebenfalls Ergebnisse, die auf einen möglichen Zusammenhang des 2D:4D zu pränatalem Testosteron hinweisen, da Studie 1 dennoch einen Geschlechtsunterschied im 2D:4D finden konnte, Studie 2 wertvolle Hinweise darauf liefert, dass Hände und Finger und auch daraus abgeleitete Längenverhältnisse geschlechtsdimorph sind, sowie Studie 3 einen Zusammenhang zwischen dem 2D:4D und geschlechtstypischem Spielverhalten von Jungen ohne Geschwister finden konnte. Der zweite Teil der Dissertation liefert zudem wichtige Anknüpfungspunkte für zukünftige Forschung zum Einfluss von mütterlichem Cortisol auf die Mutter-Kind-Interaktion und zeigte, dass Letztere womöglich negativ durch die akute HPA-Achsen-Aktivität und vermehrter Cortisolausschüttung beeinflusst wird.

Abstract

Hormones serve as molecular messengers and play a central role in regulating various physiological processes. This dissertation examines the relevance of two selected hormones, testosterone and cortisol, in closer detail. Testosterone is involved in the prenatal development of primary sexual characteristics and predisposes the development of secondary sexual characteristics. However, measuring prenatal testosterone in humans is challenging, leading research to discuss surrogate measures such as the ratio of the length of the second to fourth digit (2D:4D). Yet, the 2D:4D exhibits some weaknesses, particularly regarding a direct correlation with testosterone and its applicability in children. Cortisol is associated with the human stress response, regulated by the hypothalamic-pituitary-adrenal (HPA) axis, and can influence various cognitive processes. In the context of early mother-child interaction, cortisol may negatively affect potentially relevant cognitive resources of the mother, but, standardized experimental studies investigating the relationship between maternal cortisol and the quality of mother-child interaction are lacking.

The first part of the dissertation aimed to further investigate the applicability of 2D:4D in child samples. Three studies are presented, examining the basic assumption of a relationship between 2D:4D and testosterone levels, sex differences, which appear relatively robust in adult samples, and ultimately the applicability of 2D:4D as a measure of testosterone in relation to gender-typical behaviors of children.

Study 1 examined the relationship between testosterone levels measured in amniotic fluid and childhood 2D:4D at several postnatal time points from infancy to preschool age. Due to low testosterone levels in the amniotic fluid of female children, the relationship with 2D:4D could only be examined in the male sample, where no significant correlation was found at any of the time points. Female children consistently exhibited a larger 2D:4D compared to male children at each time point. Furthermore, the age of the children emerged as an important influencing factor, with 2D:4D increasing with age. The results suggest potential weaknesses of 2D:4D as a marker for prenatal testosterone and underscore the need for further research.

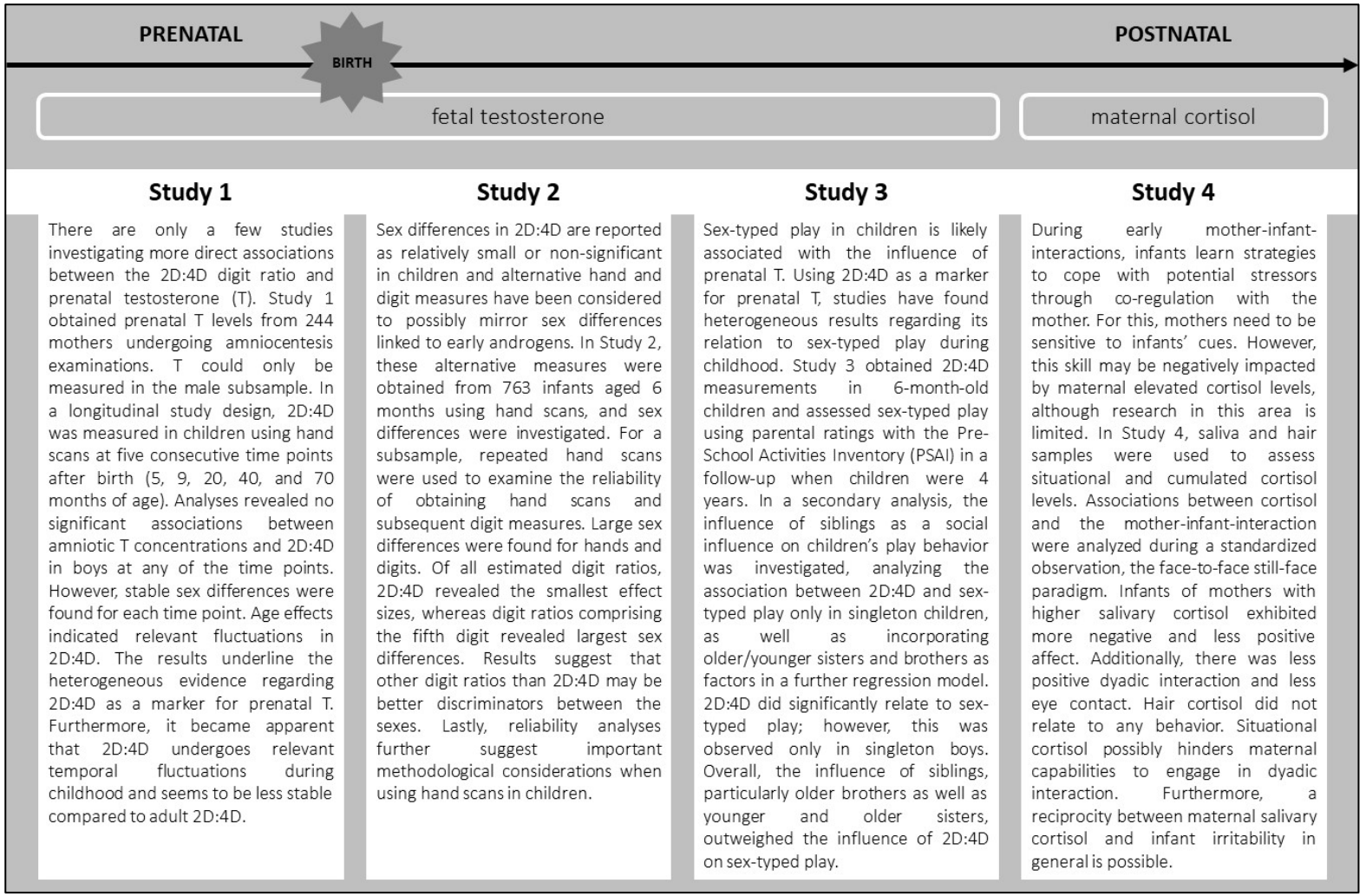
Study 2 examined gender differences in various hand and digit length ratios of children, as well as the reliability of hand scans for measuring these ratios. Male children had overall larger hands and longer digits. 2D:4D showed only small or nonsignificant sex differences, whereas digit length ratios including the fifth finger showed the largest differences. Hand scans proved to be a reliable source for measuring hand and digit length ratios; yet, the results indicated that measurement errors could occur due to lower compliance and possible distortions in scans caused by higher body fat percentage in young children compared to adults. Overall, the study supported sex differences in the hands and digits of children, but 2D:4D was the weakest measure, which may complicate its applicability especially in younger cohorts.

Study 3 investigated the relationship between 2D:4D and gender-typical play behavior in children, with the additional consideration of the influence of siblings. There was no significant difference in 2D:4D between male and female children. The relationship between 2D:4D and sex-typed play behavior was only significant in male singletons. In the overall sample, a significant influence of older and younger sisters as well as older brothers was found, but again only in male children. Sex-typical play behavior of female children was neither associated with 2D:4D nor the influence of siblings. The results suggest a potential applicability of 2D:4D as a marker for prenatal testosterone, but only for singleton boys. Further, the results suggest the importance of siblings on children's sex-typed play behavior.

The second part of the dissertation examined the relationship between maternal cortisol and the mother-infant-interaction in a standardized behavioral observation, the still-face paradigm. In Study 4, maternal cortisol levels in saliva and hair were measured, indicating situational HPA axis activity as well as long-term HPA axis activity, respectively. Salivary cortisol was associated with less positive affect, dyadic interaction and eye contact but higher infant protest behavior. Maternal hair cortisol did not appear to be associated with the mother-infant-interaction. The results suggest that acute, but not long-term, HPA axis activity may negatively influence mother-infant-interaction.

The results of the dissertation in the first part indicate relevant weaknesses of 2D:4D in its application in children. Particularly noteworthy are the lack of correlation with prenatal testosterone in Study 1 and the absence of gender differences in 2D:4D in Studies 2 and 3. However, all studies also provided support for the association between 2D:4D and prenatal testosterone to some extent, as Study 1 still found a sex difference in 2D:4D, Study 2 provides valuable insights into the sexual dimorphism of hands and fingers and derived length ratios, and, ultimately, Study 3 found a relationship between 2D:4D and sex-typed play behavior in singleton boys. The second part of the dissertation also provides important starting points for future research on the influence of maternal cortisol on the mother-infant-interaction and suggests that the latter may be negatively influenced by acute HPA axis activity and increased cortisol secretion.

Graphical overview of studies presented in the dissertation



Note: The graphical overview was created using Microsoft Office PowerPoint version 1809.

Introduction

Translated from the Greek hormãn which means “to set something in motion, stimulate or activate” (Duden online, 2023), hormones are among the most crucial signaling molecules in the human body and regulate different bodily processes, such as metabolism, homeostasis, growth or blood circulation. Hormones are signaling via feedback loops and hormonal axes and bind on specific receptors on their target organs, initiating further processes (Cole & Kramer, 2016; Kleine & Rossmanith, 2014). In addition, hormones regulate specific behaviors, like sexuality, aggression or caregiving behavior (Arnold & McCarthy, 2016; Hines, 2002). The influence of hormones on the human body is evident as early as the embryonic and fetal age(Kota et al., 2013) and also pregnancy and birth are significantly influenced by hormones (Tal & Taylor, 2021).

The current work focuses on the significance of two selected steroid hormones in early human development: the influence of testosterone (T) during early gestation on sexual differentiation and development and the influence of maternal cortisol in the context of early mother-infant-interaction. The first part of the dissertation concentrates on T, an androgen which is mainly synthesized in the gonads. It is understood to be the most important hormone for fetal development of the male phenotype and is thus significantly involved in human sexual development (Nassar & Leslie, 2022). However, the measurement of prenatal T in humans proves to be a difficult undertaking in many ways. Therefore, great efforts have been made to validate alternative measures for human prenatal T. One prominent surrogate is the ratio of the second (2D) to the fourth (4D) digit, the 2D:4D digit ratio, that has been proposed in the late 1990s (Manning et al., 1998). Since then, several studies have been conducted to prove its validity and reliability in humans. Although, more direct evidence for a link between prenatal T and 2D:4D could be derived from animal studies, in humans evidence is much weaker (see, e.g., Manning & Fink, 2023).

The second part focuses on cortisol, a glucocorticoid which is synthesized in the pituitary gland and is mostly associated with the human stress response. Cortisol is released following a stressor via the

hypothalamus-pituitary-adrenal (HPA) axis and ensures the organism's energy supply (Jänig & Baron, 2019; Lang & Föller, 2019). Further, the HPA axis follows circadian rhythms, regulating various natural activation processes in the body such as waking up (Jänig & Baron, 2019). Cortisol is immunosuppressive and is essentially for inflammatory responses in the body (Lang & Föller, 2019). The human HPA axis can be affected by long-time exposure to stress and an alteration of the HPA axis can be a risk for human physiological and psychological well-being (Agorastos & Chrousos, 2022; Chrousos, 2000), specifically, when chronic stress exposure or stressful life events occur early in life (Chaby, 2016; LeMoult et al., 2020). Efforts have been made to investigate the relation between maternal cortisol during pregnancy and its effect on fetal HPA axis (Galbally et al., 2021; Lewis et al., 2015). Evidence suggests that cortisol can also affect cognitive systems (Schwabe & Wolf, 2013) that possibly are important for maternal caregiving abilities in the early postpartum (Almanza-Sepulveda et al., 2020). Besides biological risk factors like alterations in fetal HPA axis, maternal caregiving is also a crucial factor for the early development of children and less optimal caregiving can put children at risk for later psychopathology (Calkins et al., 2013).

For both parts, it will be firstly given an overview of the theoretical background before presenting own studies that were conducted to investigate the two topics. The findings are discussed, first, within their respective contexts before giving a broader discussion and conclusion. Lastly, an outlook to future research directions is presented.

Testosterone and its influence on human sexual development

Testosterone (T) is instrumental in human sexual development and possibly contributes to the differentiation of sex-specific behaviors through its influence on brain structures and networks (Hines, 2020a). It is theorized that the release of T during critical developmental periods in humans (and other species) has enduring, so-called, organizing effects on the body and the brain (Arnold & McCarthy, 2016). In humans, these critical periods are usually associated to phases when differences in T concentrations are highest between male and female fetuses which range from 8 to 24 weeks of gestation, shortly after birth up to 6 months postnatal and again during puberty (Auyeung et al., 2013;

Hines et al., 2016; Sisk et al., 2013). With the beginning of puberty, T (along with other hormones) is released more frequently in circadian rhythms, exerting activating effects on the body, brain and behavior that are reversible (Arnold & McCarthy, 2016). This organizational-activational hypothesis, introduced by Phoenix et al. (1959) in rodents, has influenced the understanding of the influence of prenatal T in humans, although compared to rodents additional critical or sensitive periods in humans have been discussed (Sisk et al., 2013). Further, other studies explored the importance of environmental effects on shaping brain structures and networks (Arnold, 2009; Gu & Kanai, 2014; Kiesow et al., 2020). Nevertheless, the organizational-activational hypothesis is still very influential. Several studies found the early influence of prenatal T on sex-dependent behavior in animal studies (Beltz et al., 2020). Due to ethical considerations, human studies rely on indirect evidence.

Clinical samples with altered levels of androgens during gestation provide insides into effects of altered T levels during gestation. Females with congenital adrenal hyperplasia (CAH), a clinical condition with highly elevated T levels during gestation (Auer et al., 2023), show more male-typical play behavior during childhood compared to non-affected females (Beltz et al., 2023; Cohen-Bendahan et al., 2005). Further, people affected by complete androgen insensitivity syndrome (CAIS) are genotypically male and have normal levels of T, however, develop phenotypically female as T cannot bind to the specific androgen receptor and therefore has no physical effect on the body (Hughes et al., 2012). People with CAIS show more female typical play behavior during childhood compared to non-affected controls (Cohen-Bendahan et al., 2005). Research also suggests structural and functional brain differences of people affected by CAH (Swift-Gallant & Monks, 2017) and CAIS (Bakker, 2022) compared to non-affected controls. Although larger effects are reported in animal studies, T seems to have crucial organizing effects on the development of the human brain and subsequent neural networks and behavior.

In human studies, a key challenge arises from the fact that clinical samples do not reflect natural variations in T exposure, but rather clinically relevant deviations, which possibly biases the interpretation of associations between T and other variables of interest (Beltz et al., 2020). Significant

efforts have been made to evaluate other indices or measurements for prenatal T concentrations in healthy samples. One prominent, non-invasive index is the ratio of the length of the second finger to the length of the fourth finger, the 2D:4D digit ratio that will be presented in the following.

The 2D:4D digit ratio as a proxy for prenatal testosterone concentration

Manning et al. (1998) were the first to introduce 2D:4D and its proposed link to prenatal T. They proposed that higher prenatal T impacts digit length growth, particularly evident in the 2D:4D ratio, based on the involvement of homeobox (HOX) genes in both urogenital and limb development (Manning et al., 1998). Subsequent research could replicate the proposed sex difference with adult males having smaller 2D:4D compared to adult females (Hönekopp & Watson, 2010; Manning, 2002). Further, Manning and Fink (2018) noted that 2D:4D reveals reliable sex differences as early as 2 years of age, although, sex differences in 2D:4D could already be found in-utero (Galis et al., 2010; Malas et al., 2006). Animal studies supported the association between prenatal T and 2D:4D, showing that T administration during gestation affects length growth of 2D and 4D in rats (Suchonova et al., 2019; Talarovičová et al., 2009). Whereas Suchonova et al. (2019) found no effect on 2D:4D, Talarovičová et al. (2009) found a significant decrease in 2D:4D as a result of T administration during gestation. Zheng and Cohn (2011) found that androgen or estrogen receptor deletion respectively led to a feminization (larger) respectively masculinization (smaller) of 2D:4D in mice, while Swift-Gallant et al. (2021) linked larger 2D:4D to an androgen receptor overexpression. Swift-Gallant et al. (2021) argued that the direction of the sex effect in mice is reversed compared to human and that their overall finding contributes to the assumption that 2D:4D depends on androgen receptor activity. Manning and Fink (2023) argued that animal models offer evidence for a causal relationship between prenatal T and 2D:4D, Contrasting the more uncertain evidence in humans.

Comparisons between healthy controls and people with CAH or CAIS provided indirect evidence for masculinized 2D:4D in people with CAH (Brown et al., 2002; Richards et al., 2020b) and more feminized 2D:4D in people with CAIS (Berenbaum et al., 2009; van Hemmen et al., 2017). These studies lack of large samples, reveal only modest effects and 2D:4D of clinical and healthy samples largely

overlaps, although prenatal T in CAIS samples has almost no physiological effect (Brown et al., 2002; Buck et al., 2003; Nave et al., 2021; Richards et al., 2020b; Wallen, 2009). Other studies used sources to detect T levels in the fetal hormonal environment, such as maternal T levels that were measured in serum or cord blood samples. In these studies, no significant or only weak associations with infant's 2D:4D could be found (Barrett et al., 2020; Hickey et al., 2010; Richards et al., 2022; Richards et al., 2019). In fact, studies revealed that T levels measured in maternal or cord blood samples reveal only small or insignificant correlations with fetal respectively infant hormones (Penny et al., 1979; Troisi et al., 2003) and sex differences in fetal T levels are probably not reflected in maternal serum T (van de Beek et al., 2004). Amniotic fluid samples may be superior to assess prenatal sex hormone concentrations as they are collected during critical phases of highest differences between male and female fetuses' T levels (Abramovich, 1974; Finegan et al., 1989; Judd et al., 1976; Robinson et al., 1977). Indeed, Lutchmaya et al. (2004) found that the ratio of amniotic T to amniotic estradiol (E), the T/E ratio, related to the 2D:4D of 2-year-old children. Despite initial findings, subsequent studies struggled to replicate these associations (Richards et al., 2020a) or found associations between amniotic T and 2D:4D only in girls (Ventura et al., 2013). These data rely on cross-sectional studies in young children. However, it may be additionally important to collect longitudinal data, specifically when 2D:4D of children is examined, as the literature suggests significant age effects in 2D:4D (Eichler et al., 2023; Ernsten et al., 2023; McIntyre et al., 2006; McIntyre et al., 2005; Richards et al., 2017; Trivers et al., 2006; Wong & Hines, 2016), against the claim of Manning and Fink (2018) that 2D:4D should be independent of age. Further, 2D:4D sex differences in children are often insignificant or only small (Ernsten et al., 2021; Knickmeyer et al., 2011; McIntyre et al., 2005; Trivers et al., 2006). As the potential influence of activating effects via T with onset of puberty and subsequent effects on body development (Delemarre-van de Waal et al., 2001) cannot be ruled out in older samples, it may be specifically appealing to find consistent and large sex differences in 2D:4D of prepubertal cohorts as these are possibly less influenced by effects of fluctuating T.

Finally, indirect evidence suggests an association between 2D:4D and various behaviors or characteristics likely influenced by prenatal T, for example, children's play behavior (Hönekopp & Thierfelder, 2009; Körner et al., 2020; Mitsui et al., 2016; Wong & Hines, 2016), aggression (Bailey & Hurd, 2005; Hönekopp & Watson, 2011), physical fitness (Pasanen et al., 2021) or conditions like autism (Hönekopp, 2012; Teatero & Netley, 2013). Human behavior generally exhibits smaller sex differences compared to animal behavior (Hines, 2020b), possibly influenced by psychological, social and environmental factors (Hines, 2015). Nevertheless, behavior like sex-typed play in children shows large sex differences (Cohen's d = 2.8; Hines, 2020b), can be already observed during the first year of life (Todd et al., 2018) and, as sex-typed play can also be observed in non-human primates, a biological component seems reasonable (Lonsdorf, 2017; Marley et al., 2022). Different studies reported significant associations between children's sex-typed play with amniotic T as well as T measured in maternal serum during pregnancy (Auyeung et al., 2009; Hines et al., 2002), whereas others did not find such an association (Knickmeyer et al., 2005; Spencer et al., 2021; van de Beek et al., 2009). Further, many studies found small to moderate associations between sex-typed play and 2D:4D as well (Hönekopp & Thierfelder, 2009; Körner et al., 2020; Mitsui et al., 2016; Wong & Hines, 2016), whereas others did not (Barrett et al., 2020).

In sum, evidence suggests a link between 2D:4D prenatal T levels. Yet, in the few available studies that attempted to examine this relationship more directly, considerable heterogeneity remains in the study results (Lutchmaya et al., 2004; Richards et al., 2020a; Ventura et al., 2013). As a meta-analysis suggests, sex differences in 2D:4D have to generally be considered as small to moderate (Hönekopp & Watson, 2010). As one of the most cited studies had only a small sample (N = 29, see, Lutchmaya et al., 2004), it cannot be ruled out that the finding of true positive results was affected. Further, most of the studies that did find robust sex differences related to adult samples (Hönekopp & Watson, 2010), whereas in younger samples age seems to be an important factor that should be considered (Ernsten et al., 2021; Knickmeyer et al., 2011; McIntyre et al., 2005; Trivers et al., 2006). Meaningful longitudinal data in young cohorts report significant changes in 2D:4D dependent on age

(Knickmeyer et al., 2011; McIntyre et al., 2006; McIntyre et al., 2005). Lastly, alternative hand and digit measures or ratios have been discussed to reflect sex differences, with some studies reporting better applicability of other digit ratios, specifically, in young children (McIntyre et al., 2006). These issues have so far received insufficient attention in research, although, important considerations regarding the temporal stability and applicability of 2D:4D result as a consequence.

The relevance of T for sexual development and the endeavor to find a possibility to measure prenatal T and investigate subsequent effects on the developing human have been described above. Besides T, there are other relevant hormones that potentially affect the early development of humans. The current work will further discuss the importance of cortisol in the context of the early mother-infant-interaction as well as socio-emotional development of infants.

Cortisol in the context of early mother-infant-interaction

In response to a potential stressor, the body prepares for a flight or fight reaction. The energy supply for this reaction is assured, amongst others (for the relevance of the sympathetic nervous system, see, Birbaumer & Schmidt, 2010b, which will not be discussed further in this work), through activation of the HPA axis (Cole & Kramer, 2016). Upon encountering a stressor, the hypothalamus is activated and produces corticotropin releasing hormone (CRH), which stimulates the anterior pituitary gland and induces the production of adrenocorticotropic hormone (ACTH; Cole & Kramer, 2016). Subsequently, the glucocorticoid cortisol is synthesized in the adrenal cortex and ensures gluconeogenesis (Cole & Kramer, 2016). HPA axis activity is regulated by a negative feedback loop, wherein cortisol release suppresses the production of CRH and ACTH when the stressor resolves, which ultimately results in homeostasis (Birbaumer & Schmidt, 2010a). If the stressor persists, which is commonly the case facing chronic psychosocial stressors (Birbaumer & Schmidt, 2010b), long-time changes in HPA axis activity and cortisol levels due to elevated HPA axis activity can be the consequence (Miller et al., 2007).

Women transitioning to motherhood face crucial changes on various levels, including physical, psychological, emotional and social (Emmanuel & St John, 2010; Nelson, 2003), that can serve as

stressors (Emmanuel & St John, 2010). The potential consequences of acute or chronic maladaptation of the maternal stress response during the early postpartum period can put mothers (Mughal et al., 2018; Saxbe et al., 2018) and their infants (Hruschak et al., 2022; Richter et al., 2022; Van den Bergh et al., 2020; van den Heuvel, 2022) at risk. As infants cannot yet explain their needs, mothers need to adequately perceive, interpret and react to their infant's cues in a timely and qualitative manner, a skill that has been described as maternal sensitivity (Ainsworth et al., 1978; Bowlby & Ainsworth, 2000; DiCorcia & Tronick, 2011; Shin et al., 2008). Maternal sensitivity has been described as a "dynamic process involving maternal abilities, reciprocal give-and-take with the infant, contingency on the infant's behavior and quality of maternal behaviours" (Shin et al., 2008, p. 306). It has already been shown that maternal sensitivity led to fewer socio-emotional and behavior problems (Behrendt et al., 2019; Frick et al., 2018) and better stress regulation (Blair et al., 2006; Provenzi et al., 2017) in infants, whereas maternal insensitivity could be associated with elevated infant stress responses (Enlow et al., 2014). Further, infants of sensitive mothers tended to be more securely attached (Bowlby & Ainsworth, 2000).

Especially during mother-infant-interactions, along with interactions with other important caregivers (considering the importance of fathers, see Bakel & Hall, 2020; Bakermans-Kranenburg et al., 2019), maternal sensitivity is a core ability for mothers. As infants cannot yet fully adapt and cope with different stimuli (Schore, 2001), mothers serve in a coregulating function, supporting their infants to cope with potential stressors (Laurent et al., 2016). According to the mutual regulation model (Gianino & Tronick, 1988; Tronick & Gianino, 1986), the interaction between mothers and infants follows phases of matches and mismatches. Matching states describe interactive patterns where „there is a fittedness of maternal regulatory input to infants' signaled regulatory needs" (DiCorcia & Tronick, 2011, p. 1596). When dyadic interactions fail to achieve this „fittedness" (DiCorcia & Tronick, 2011, p. 1596), a transition to a mismatch state follows. Phases of mismatches commonly occur in the early mother-infant-interaction, yet, it is necessary that the dyad experience adequate reparation of the interaction and a change back to a matching state (DiCorcia & Tronick, 2011; Gianino & Tronick, 1988). Reparation of mismatching states foster infants' capabilities to cope with future interactional mismatches and lead to

the development of infants' self-regulatory behavior, whereas chronically ineffective reparation may leave the infant stressed and thus hinders the socio-emotional and cognitive development (DiCorcia & Tronick, 2011).

Elevated cortisol may be a risk-factor for the mother-infant-interaction and effective reparation after mismatching states. It has been shown that cortisol leads to more habitual rather than sensitive behavior due to a shift in memory systems (Schwabe & Wolf, 2013). As the authors note, behavior based on more habitual memory systems indeed allows the individuum to retain resources that are important to cope with the stressor, but it can also leave the individuum less flexible in their behavior (Schwabe & Wolf, 2013). This shift potentially interferes with maternal capabilities to engage in sensitive behavior towards her infant. However, the role of cortisol for maternal sensitivity has not been extensively investigated. Thompson and Trevathan (2008) found that decreasing maternal cortisol, interpreted by the authors as lower maternal physiological arousal due to the laboratory examination, was related (indexed by a statistical trend) to greater levels of mother-infant-synchrony during a 10-minute free-play observation with 3-month-old children. The authors used salivary cortisol concentrations (SCC) as a marker for stress (Thompson & Trevathan, 2008), which is a commonly used tool for the assessment of the situational HPA axis activity or response (Kirschbaum & Hellhammer, 2008). Mueller et al. (2021) used an experimental manipulation (i.e., infant crying) for mothers that should induce stress. They found that the interaction between mothers in the stressed group and their 4-month-old infants had to be terminated more often due to higher infant distress compared to mothers in the control group who did not experience the stress manipulation (Mueller et al., 2021). These authors did not control for an increase in maternal cortisol and the focus of the study was on infant distress (Mueller et al., 2021). Nonetheless, the results of Thompson and Trevathan (2008) and Mueller et al. (2021) suggest that the mother-infant-interaction could be impaired by higher cortisol levels and that infants of highly stressed mothers may also experience higher amounts of stress and are more irritable or dysregulated.

Besides situational effects of cortisol, long-term activation of the HPA axis due to persisting stressors can be associated with several psychiatric disorders and mental health problems (Guidi et al.,

2020; Lupien et al., 2009; Marin et al., 2011). If a stressor occurs and cannot be resolved, the body potentially fails to achieve homeostasis and induces allostasis to adapt to the demands of the ongoing stressor (Juster et al., 2010). This allostasis can entail a collapse of different bodily systems if it becomes chronic (Juster et al., 2010). Heightened cortisol levels over a longer period of time could indicate chronic stress exposure (Juster et al., 2010) and the analysis of cortisol levels in human hair has been discussed to reflect this long-term HPA axis activity (Meyer & Novak, 2012). In the context of mother-infant-interaction, Tarullo et al. (2017) already reported less positive and more negative interaction patterns in a free-play related to higher maternal hair cortisol concentrations (HCC) in mother-infant-dyads at the age of 6 months. Further, Khoury et al. (2020) found that higher HCC and maternal depressive symptoms during pregnancy and 4 months after birth were predictive for more negative and disrupted mother-infant-interaction in a standardized behavioral observation at 4 months after birth. Higher HCC in the group of mothers scoring low in depression related to more intrusive behavior during the mother-infant-interaction, whereas mothers with higher HCC and higher depressive symptoms showed more withdrawal behavior during the observation (Khoury et al., 2020). Therefore, both situational and long-term cortisol can potentially be risk factors for the mother-infant-interaction.

In sum, stress as a daily occurrence potentially affects most families (Crnic & Greenberg, 1990; Suhr, 2019) and can be an important risk factor for maternal sensitivity in the context of mother-infant-interaction possibly due to elevated cortisol levels (Thompson & Trevathan, 2008) and prolonged maternal mental health problems (Duthie & Reynolds, 2013; Öhman et al., 2007). However, there is a general lack of studies investigating situational and more long-term effects of cortisol on the mother-infant-interaction.

Research questions

The current work comprises research on two relevant hormones that shape the early development of infants and the early mother-infant-interaction. Testosterone (T) has an effect on sexual and gender development already in-utero (Arnold & McCarthy, 2016). The measurement of prenatal T in humans is difficult and researchers are focusing on proxies that potentially reflect early influences of

prenatal T. The 2D:4D digit ratio is a commonly used marker for prenatal T exposure (Manning et al., 1998), however, research on its validity (Lutchmaya et al., 2004; Richards et al., 2020a) and reliability (Eichler et al., 2023; Knickmeyer et al., 2011; McIntyre et al., 2006; McIntyre et al., 2005; Richards et al., 2017; Trivers et al., 2006; Wong & Hines, 2016) is heterogeneous, specifically, in younger cohorts. The first aim of the current work was to investigate the usefulness of the 2D:4D digit ratio as a proxy for prenatal T exposure in cohorts ranging from 5 to 70 months of age. In a first study, the association between 2D:4D and amniotic T measured during the second trimester was investigated in a longitudinal study design with an additional focus on age effects in 2D:4D. Study 2 addressed sex differences in 2D:4D and other sex-dimorphic indices of the human hand and human digits that have also been proposed to reflect prenatal T exposure and the reliability of obtaining such measurements. Study 3 intended to investigate the association of 2D:4D and sex-typed play behavior. In sum, these studies allow conclusions about the hypothesized usability and applicability of 2D:4D as a marker of prenatal T in young children.

Secondly, the current work focused on the importance of cortisol in the context of early mother-infant-interaction. Cortisol in the context of stress potentially impairs maternal capabilities to engage in sensitive caregiving (Mueller et al., 2021; Schwabe & Wolf, 2013; Thompson & Trevathan, 2008), although, meaningful studies are currently lacking. The aim of Study 4 was to investigate the association of maternal momentary and cumulated cortisol levels and the mother-infant-interaction. It is assumed that elevated maternal cortisol levels relate to less positive and more negative interactional patterns, For the investigation, a standardized laboratory paradigm, the elaborated face-to-face still-face (FFSF) paradigm (Tronick et al., 1978), was used. Further, maternal cortisol levels were assessed by obtaining saliva samples which reflected momentary cortisol levels of mothers at the beginning of the experiment, as well as hair samples which reflected cumulated cortisol levels during the past 3 months. The results may provide valuable endpoints that can be used for future research on the complex effects of maternal cortisol during the early postpartum period.

Overview of studies

Partial results of the presented work have been published or submitted: Study 1 is based on Ernsten et al. (2023), Study 2 on Ernsten et al. (2021), Study 3 on Ernsten et al. (2024) and Study 4 is based on Ernsten et al. (accepted for publication after minor revision) . All articles and manuscripts can be found in the appendix.

Study 1:

Only three previously publishes studies investigated associations between prenatal T concentrations and 2D:4D using amniotic fluid samples which are probably best suited to reflect fetal T levels. In Study 1, amniotic fluid samples of 244 pregnant women were obtained during the second trimester of gestation. The aim was to evaluate the association between amniotic T and 2D:4D of children in a longitudinal study design addressing inconsistencies in previous findings. Children's 2D:4D was obtained at five consecutive measurement timepoints from infancy to pre-school age (5, 9, 20, 40 and 70 months of age) using hand scans. In line with the underlying hypothesis, a negative association between amniotic T and children's 2D:4D was assumed. Further, sex differences (male 2D:4D < female 2D:4D) and the influence of age on 2D:4D were further investigated.

Study 2:

Relatively robust sex differences are reported for adult 2D:4D, whereas in children, results are more heterogenous and age seems to be an important influencing factor. Further, other hand and digit measures were already reported to obtain sex differences, but the measurement differed significantly and most of the measures were evaluated in adults. Therefore, the aim of Study 2 was to compare different hand and digit measures in terms of sex differences using the same measurement technique to increase comparability between measures, specifically in children. Further, as one of the first studies, the reliability of hand scans in very young children was investigated. Different hand and digit measures were derived via hand scans of 763 children aged 6 months. For a subsample, hand scans were obtained twice and a reliability analysis was conducted.

Study 3:

Children's play behavior reveals large sex differences and supposedly relates to the influence of prenatal T. Some studies already reported significant associations between children's play behavior and 2D:4D as a proxy for prenatal T, although leaving behind a considerable heterogeneity. Further, sex-typed play is known to be affected by other influences such as social influences due to siblings. In Study 3, the aim was to investigate the association between 2D:4D and sex-typed play of children and it was assumed that smaller 2D:4D related to more male-typed play behavior. Further, the association was tested only in singleton children as these are assumingly less influenced by social factors that may relate to sex-typed play. Lastly, the association between sex-typed play, 2D:4D and older or younger sisters or brothers was investigated. 2D:4D was obtained via hand scans when children were 6 months of age and sex-typed play behavior was assessed in a follow-up 3.5 years later through parent's reports.

Study 4:

The early mother-infant-interaction is a crucial experience for the developing infant and promotes infant's development of self-regulation while the mother serves in a co-regulating function. For this, the mothers need to be sensitive to her infant's cues and needs which can potentially be impaired by elevated maternal cortisol levels. Evidence on a negative effect of maternal cortisol on the quality of mother-infant-interaction in standardized assessments is scare. Therefore, the aim of Study 4 was to investigate the association of cortisol using two endocrine markers for cumulated HPA axis activity (i.e., hair cortisol concentrations) and for momentary HPA axis activity (i.e., salivary cortisol concentrations) and the mother-infant-interaction following a standardized paradigm, the face-to-face still-face (FFSF) paradigm. The FFSF promotes a sudden break in dyadic interactive behavior and, during the last phase of the paradigm, dyadic regulation can be observed. It was assumed that the amount of maternal and infant positive affect as well as dyadic behavior during the last phase of the FFSF is negatively associated with maternal cortisol, while infant negative affect is positively associated with maternal cortisol levels.

Study 1

A longitudinal study of prenatal amniotic sex hormones and digit ratio (2D:4D) in children aged 5 to 70 months

(Ernsten et al., 2023, PLOS ONE)

Amniocentesis examinations seem appropriate to investigate more direct associations between fetal T concentrations and postnatal 2D:4D (van de Beek et al., 2004). Amniocentesis examinations, typically conducted in the second trimester, coincide with a period of significant T concentration differences between male and female fetuses (Abramovich, 1974; Finegan et al., 1989; Judd et al., 1976; Robinson et al., 1977). However, limited studies have investigated the link between amniotic T and 2D:4D. Lutchmaya et al. (2004) found a significant association between the amniotic T/E ratio measured during the second trimester with 2D:4D of 29 2-year-old children. It is probably the most cited study and is often referenced as evidence for the link between prenatal T and 2D:4D, although the sample was considerably small (N = 29) and the authors solely controlled for sex in their multiple regression model rather than investigating the association separated by sex, which should always be done when investigating variables that both exhibit sex differences (as recommended by, e.g., Breedlove, 2010; Constantinescu & Hines, 2012). Richards et al. (2020a) attempted to replicate the findings of Lutchmaya et al. (2004), but found no significant association between amniotic T respectively E and 2D:4D in a sample of 66 children aged 4.5 years. Ventura et al. (2013) sampled amniotic T during the second trimester of gestation and obtained 2D:4D in newborn infants on average 38.9h after birth. Smaller 2D:4D significantly related to higher amniotic T levels, although only in female newborns (Ventura et al., 2013). Interestingly, Ventura et al. (2013) were the only ones to detect a significant sex difference in 2D:4D in the left hand of newborns, whereas Lutchmaya et al. (2004) and Richards et al. (2020a) did not observe a significant sex difference in 2D:4D. Notably, all three studies obtained 2D:4D at different ages, even though age has been discussed as a potentially relevant factor, specifically, for children's 2D:4D (Knickmeyer et al., 2011; McIntyre et al., 2006; McIntyre et al., 2005). Thus, Study 1 aimed to investigate

the association between amniotic T and 2D:4D, sex differences in 2D:4D and age effects of 2D:4D in a longitudinal study design.

Methods

The study comprised data of 244 mothers who underwent amniocentesis examination between gestational week 14 to 18. The total concentration of T (in ng/ml) as well as the T/E ratio was of interest, but, as 97.4% of female infant's T concentrations were between or underneath the limit of quantification (0.05 ng/ml) respectively the limit of detection (0.02 ng/ml), only male infant's T concentrations could be used for data analysis. In five consecutive follow-up sessions at ages 5, 9, 20, 40 and 70 months of age, 2D:4D of children was assessed using indirect measurements by scanning the ventral surface of the hand with an image scanner and the computer program *Autometric* (DeBruine, 2004) for measurement of digit lengths and calculation of 2D:4D. Furthermore, the difference between right and left 2D:4D ($D_{[r-l]}$) was additionally calculated. This measure was introduced by Manning (2002) and hypothesized to reflect differential influences of hormones on the right and left side of the body. It is assumed that effects of T could be more pronounced in the right hand which could already been underpinned by a meta-analysis of Hönekopp and Watson (2010) who reported larger sex differences in right 2D:4D. The association between amniotic T as well as the association of the T/E ratio with 2D:4D to five measurement timepoints in boys was analyzed using multilevel linear regression models entering amniotic T respectively the T/E ratio as predictors into the model. To account for potential effects of age and timepoint of amniotic fluid sampling, these factors were also entered as predictors. 2D:4D was the criterion. The same analyses were carried out for $D_{[r-l]}$. Further, to assess differential effects of age, sex and hand, multilevel linear regression models were conducted using the predictors age, sex (coded as 0 = boys, 1 = girls) and hand (coded as 0=left, 1=right) on 2D:4D respectively $D_{[r-l]}$. Lastly, correlational analyses of 2D:4D between the consecutive measurement timepoints were conducted as an indicator for temporal stability of 2D:4D.

Results and discussion

No significant association could be found for boys' amniotic T as well as the T/E ratio for both 2D:4D and $D_{[r-l]}$ for none of the five consecutive measurement timepoints (all $p \geq .304$). Next, analysis revealed significant age ($p < .001$) and sex differences ($p < .001$) in 2D:4D. For sex, larger 2D:4D for girls could be found for all measurement timepoints (all $p \leq .022$). 2D:4D decreased from 5 to 9 months ($p = .089$) and increased from 20 to 40 months ($p < .001$) as well as from 40 to 70 months ($p < .001$), indicating age dependent fluctuations. Between 9 and 20 months, there were no significant differences in 2D:4D ($p = .655$). Significant age and sex ($p = .050$) as well as age, sex and hand interactions ($p = .036$) indicated differential fluctuations of 2D:4D over time that varied depending on the sex and the hand of children. Correlations between measurement timepoints were small to moderate from 5 to 40 months of age, with r ranging between .29 and .50 (all $p \leq .037$) and moderate to large from 40 to 70 months of age, with r ranging between .61 and .66 (all $p < .001$), indicating higher temporal stability and reliability with increasing age.

The current results add to the heterogeneous and scarce findings on the association between amniotic T and 2D:4D. The finding of a negative association between amniotic T or the T/E ratio as indicated by Lutchmaya et al. (2004) could not be replicated, which aligns with the replication of Richards et al. (2020a) who also found no significant association. Unfavorably, the finding of Ventura et al. (2013), who found a significant association between 2D:4D and T in girls only, could not be replicated in the current sample as T levels of girls could not be measured. Ventura et al. (2013) hypothesized a potential threshold for the influence of T on the developing fetus. They suggested that only girls are affected by values above this threshold as boys are generally exposed to much higher levels of T during gestation (Ventura et al., 2013). Lutchmaya et al. (2004) found the association for both sexes. Although amniotic fluid samples are potentially the best source for investigating more direct associations between fetal hormone concentrations and 2D:4D in humans (van de Beek et al., 2004), measurements can be biased. Amniocenteses are carried out as part of clinical diagnostics and their timing is therefore not necessarily based on when the greatest differences between male and female fetuses in terms of

intrauterine T concentrations occur, even if the two periods can overlap (Kuijper et al., 2013, reports greatest differences in T between gestational week 14 and 22). The available literature reports diverging measurement time points (Lutchmaya et al. (2004) reported that amniocenteses were carried out during the second trimester without further specification, Ventura et al. (2013) reported examinations on average during gestational week 17 and Richards et al. (2020a) reported a range of gestational week 15 to 22 and lastly, in the current study, amniocenteses took place between gestational week 14 to 18), which could be a significant source of variance and possibly impairs the comparability of studies. As amniocenteses are single measurements they possibly lack in reliability and repeated measurements in humans are not possible due to obvious ethical considerations. Nevertheless, evidence from animal studies promote the assumption that 2D:4D or digit growth in general relates to prenatal androgens and that a direct manipulation of prenatal T levels results in changes in digit lengths (Suchonova et al., 2019) or 2D:4D (Talarovičová et al., 2009). Other studies, argue against the notion that 2D:4D correlates with total T levels. Instead, they propose that it serves as a measure of androgen receptor sensitivity to T (Swift-Gallant et al., 2021; Zheng & Cohn, 2011), which further questions the applicability of total T levels measured in amniotic fluid. Studies also investigated the association between prenatal T measured in amniotic fluid and other sexually dimorphic traits in humans and reported significant study results (see, e.g., Auyeung et al., 2009; Auyeung et al., 2013; Baron-Cohen et al., 2015), which could be interpreted as evidence for the applicability of amniocenteses as a valid source for the measurement of amniotic T.

In sum, it seems appropriate to investigate T levels measured in amniotic fluid, although, methodological considerations have to be taken into account. The non-significant association between amniotic T and 2D:4D in the current study and heterogeneous study results reported in other studies (Lutchmaya et al., 2004; Richards et al., 2020a; Ventura et al., 2013) leaves behind considerable uncertainty regarding the validity of 2D:4D as a measure of prenatal T. Another source of variance that affects 2D:4D possibly has to be considered and was not addressed in the current study.

Age, sex and hand dependent variations in 2D:4D further promote the idea of other influencing factors and weaken the reliability of measuring 2D:4D. Sex differences were evident for every time point but age effects revealed relevant fluctuations in 2D:4D, which is in accordance with other studies (McIntyre et al., 2006; McIntyre et al., 2005; Trivers et al., 2006). Further, stability in 2D:4D seemed to increase with age, which could align with other studies reporting higher stability in 2D:4D of older study group populations (Hönekopp & Watson, 2010). Age dependent fluctuations in 2D:4D possibly result from methodological problems in obtaining digit measurements in very young cohorts. Young children have a higher amount of body fat tissue (Kuzawa, 1998), which possibly affects the reliable assessment of landmarks on the ventral surface of the hand, specifically, when using indirect measurements like the current study. Studies with very young children possibly also face more problems in following standardized measurement protocols due to less cooperation in very young children compared to adults (e.g., holding still while obtaining a hand scan). Although, sex differences in 2D:4D were evident in the current study, in younger cohorts, sex differences are generally reported as small, if they are found at all (Knickmeyer et al., 2011; McIntyre, 2006; McIntyre et al., 2006; McIntyre et al., 2005; Trivers et al., 2006). The investigation of 2D:4D in cohorts that have not yet reached puberty seems appealing, as it rules out a possible confounding due to effects of circulating sex hormones which also promote growth (Delemarre-van de Waal et al., 2001). As there is a large time-frame from approximately 6 months of age until the onset of puberty when sex hormones do not vary to a greater extent between female and male children (Auyeung et al., 2013; Lamminmaki et al., 2012), it seems reasonable to conduct research on the association between prenatal T and 2D:4D during this time-frame. Nevertheless, evidence in prepubertal cohorts does not suggest stable sex differences in 2D:4D (Knickmeyer et al., 2011; McIntyre, 2006; McIntyre et al., 2006; McIntyre et al., 2005; Trivers et al., 2006), although, prenatal T levels largely differ between male and female fetuses (Robinson et al., 1977; van de Beek et al., 2004). In sum, research in younger cohorts only provides little evidence for the assumption that 2D:4D reflects prenatal T action and brings up important methodological implications, specifically considering the influence of age on children's 2D:4D.

Conclusion

The results of the longitudinal study highlight the complexity of the association between prenatal T obtained in amniotic fluid and 2D:4D, emphasizing the need for further research to clarify this relationship in human studies. No significant association between amniotic T and 2D:4D in boys, but significant sex and age effects could be found. In light of the heterogeneous and scarce research data, it may not be justified to use 2D:4D as a proxy for prenatal T, underscoring the need for cautious interpretation and further investigation into the direct associations between prenatal T and 2D:4D in human studies.

Study 2

Investigating the reliability and sex differences of digit lengths, ratios, and hand measures in infants

(Ernsten et al., 2021, Scientific Reports)

Study 1 indicated some uncertainty concerning 2D:4D's association with prenatal T, however, a vast amount of studies refers to relatively stable sex differences in adult 2D:4D (male < female; Hönekopp & Watson, 2010) as an indicator for its applicability as a marker for prenatal T. Inconsistencies arise while looking at studies on children's 2D:4D, as many studies only found small or nonsignificant sex differences (Knickmeyer et al., 2011; McIntyre et al., 2006; Trivers et al., 2006), although Study 1 contributed to the finding of significant sex differences in young children. Some researchers explored alternative measures to obtain stable sex differences in children's hands and digits, suggesting that androgen action influences hands and digits more broadly, not just 2D:4D (Goodman, 2002; McIntyre, 2006).

Previous research has examined various digit ratios, including second (2D), third (3D), fourth (4D) and fifth (5D) digit and reported sex differences in several studies (Dressler & Voracek, 2011; Kumar et al., 2017; Kyriakidis, 2021; Kyriakidis & Papaioannidou, 2008; McFadden & Shubel, 2002). Some studies even reported larger effects in other digit ratios than 2D:4D (Dressler & Voracek, 2011; Kumar et al., 2017; McIntyre et al., 2005), although, they varied in their measurement techniques potentially affecting comparability. Other studies were interested in the length of one digit relative to the length of all digits taken together (Loehlin et al., 2009) and it could be shown that females exhibited longer relative lengths of 2D (rel2) and 3D (rel3) while males exhibited longer relative lengths of 4D (rel4) and 5D (rel5) (Loehlin et al., 2009; Stenstrom et al., 2011). These studies solely relied on adult samples. To investigate assumed larger effects of T on the right versus left body side, Manning (2002) introduced $D_{[r-l]}$ (i.e., right 2D:4D minus left 2D:4D). Likewise to relative digit lengths, this measure has not been

extensively investigated in younger cohorts. Richards et al. (2019) investigated sex differences in $D_{[r-l]}$ in infants with non-significant results. The hand itself is also used as a discriminator between the sexes (e.g., in forensic contexts, see Kanchan & Krishan, 2011). Therefore, the hand length, width and the hand index (i.e., the hand length divided by the hand width multiplied by 100) are often used for the investigation of sex differences (Aboul-Hagag et al., 2011; Kanchan & Krishan, 2011; Krishan et al., 2011). Yet, no sex differences were found in aborted fetuses (Malas et al., 2006; Raziye et al., 2016).

In sum, the literature suggests alternative measures derived from digit and hand measurements as potential substitutes for 2D:4D in investigating sex differences. Evidence lacks particularly in young cohorts and variations in measurement methods across studies highlight the need for comprehensive comparative research. Study 2 aimed to address this gap and investigated sex differences in a range of hand and digit measures and also incorporated reliability analyses.

Methods

Data was obtained in a sample of *N* = 1381 6-month-old children, who were on average 195.1 days old (*SD* = 8.4). Hand scans of the right and left hand were gathered using a flatbed scanner and the computer program *Autometric* (DeBruine, 2004). As only full samples of hand scans (i.e., valid scan of the right and left hand) were included in data analysis and some of the scans could not be measured precisely (e.g., due to masking of relevant landmarks or digits that were not straight), the sample decreased to *n* = 763 (52.3% male). Of these, another subsample of *n* = 180 (56% male) hand scans were eligible for the analysis of hand lengths and widths, as relevant landmarks were covered by infants' clothes in a large number of scans. Another subset of *n* = 130 infants of the total sample participated in a repeated measurement of hand scans at the end of the experiment, which were used for reliability analyses in the current study. Of these 130 repeated scans, only *n* = 19 allowed the measurement of hand lengths and widths. Therefore, the reliability analysis was restricted to digit measures.

A precise description of the calculation of each measure can be found in Table 1. Of interest were different groups of measures, namely the length of digits (i.e., 2D, 3D, 4D, 5D), digit ratios (i.e., 2D:3D, 2D:4D, 3D:4D, 3D:5D, 4D:5D), relative digit lengths (i.e., rel2, rel3, rel4, rel5), the difference

between right and left 2D:4D ($D_{[r-l]}$), as well as the hand length, width and the hand index (i.e., the hand length divided by the hand width multiplied by 100). All hand and digit measures were calculated using the respective landmarks (see Table 1) on the ventral surface of the hand and lengths were given in pixels (100 pixels refer to 2.54 cm using a 100 dpi monitor). All scans were rated twice by two independent raters with a high inter-rater-reliability for digit lengths (ICC = .94-.97, all p <.001) and for hand lengths and widths (ICC = .86-.96, all p <.001), supporting the use of indirect measurements via hand scans (Kemper & Schwerdtfeger, 2009; Mikac et al., 2016). Total lengths were accordingly averaged from both ratings. Correlational analyses of repeated scans were done to test for the reliability of digit measures. To test for sex and hand differences, multiple mixed 2x2x*n* ANOVA (*n* for different factor levels of groups of measures, see Table 1) investigated the main effects and interaction of the between-subjects factor sex (male versus female) as well as the within-subject factors hand (right versus left) and the according group of digit measure (e.g., digit length with four factor levels 2D, 3D, 4D and 5D). For all 2x2x*n* ANOVA, post-hoc analyses were carried out for each hand separately using 2x*n* ANOVA with sex as between-subjects factor and group of digit measure as within-subject factor. Post-hoc independent samples *t* tests were used to disentangle effects of sex in each factor level of the respective group of digit measure. Mixed 2x2 ANOVA with sex as between-subjects factor and hand as within-subject factor were carried out for hand length, width and index, separately. Post-hoc *t* tests were carried out to investigate the effect of the independent variable sex on hand length, width and index. Sex differences in $D_{[r-l]}$ were tested using independent samples *t* tests. Cohen's *d* is reported as the effect size and interpreted according to Cohen (1988): $d \leq 0.20$ small, $d \leq 0.50$ medium and $d \leq 0.80$ large effect.

Table 1

Description of Each Measure With the Additional Factor Levels.

Measure	Factor levels	Factor	Description
Digit length	4	2D	Length of the second digit. [a]
		3D	Length of the third digit. [a]
		4D	Length of the fourth digit. [a]
		5D	Length of the fifth digit. [a]
Digit ratio	5	2D:4D	Ratio of the second to the fourth digit.
		3D:4D	Ratio of the third to the fourth digit.
		2D:5D	Ratio of the second to the fifth digit.
		3D:5D	Ratio of the third to the fifth digit.
		4D:5D	Ratio of the fourth to the fifth digit.
Relative digit length	4	rel2	Ratio of the second digit to the cumulated length of all four digits.
		rel3	Ratio of the third digit to the cumulated length of all four digits.
		rel4	Ratio of the fourth digit to the cumulated length of all four digits.
		rel5	Ratio of the fifth digit to the cumulated length of all four digits.
Right-left asymmetry in 2D:4D	1	$D_{[r-l]}$	Right 2D:4D minus left 2D:4D
Hand length	1	Length	Length of the hand. [b]
Hand width	1	Width	Width of the hand. [c]
Hand index	1	Index	Ratio of the hand worth to the hand length multiplied by 100.

Note: [a] Measured from the proximal tip of the according digit to the midpoint of the distal flexion crease of the according digit; [b] Measured from the tip of the third digit to the midpoint of the central crease of the carpus; [c] Measured from the ventral crease above the head of the second metacarpal bone to the central crease of the head of the fifth metacarpal bone (Table 1 adapted from Ernsten et al., 2021).

Results and discussion

The main aim of study 2 was the investigation of sex differences in the different hand and digit measures with an additional focus on the reliability of obtaining such measurements in a large cohort of 6-month-old infants. Here, only effects of sex and the reliability analysis are further discussed.

Correlational analyses between repeated measurements indicated moderate to high reliability, with large correlations found for digit lengths (r = .68 to .94, all p <.001) and moderate to large correlations for ratios and indices, (r = .33 to .75, all $p \leq .003$). Other studies also indicated that repeated scans do not correlate perfectly (Mikac et al., 2016), indicating a source of measurement error. Mikac

et al. (2016) recommend standardized measurement protocols to enhance precision. In infant studies, challenges probably exist due to less cooperation and understanding (e.g., holding still, straightening of digits while pressing onto the scanner glass) compared to adult samples. Another source of measurement error may be infants' higher amount of body fat tissue (Kuzawa, 1998), possibly leading to a shifting of important landmarks used to measure digit and hand lengths. Reliability analyses are strongly indicated for future studies in young children.

A sexual dimorphism was found in hands, digits and derived ratios of 6-month-old infants. Compared to girls, boys had larger hands and digits with moderate effect sizes in digit lengths (all p <.001, d = 0.52 to 0.75), hand length, (all $p \leq .016$, d = 0.37 to 0.52) and hand width (all $p \leq .001$, d = 0.50 to 0.59). The hand index differed only to a small extent between girls and boys (all $p \geq .134$, d = 0.18 to 0.23). While sex differences in adult hands seem established (Kanchan & Krishan, 2011), in younger, prepubertal cohorts it seems less evident (e.g., no significant sex difference in aborted fetuses, see Malas et al., 2006; Raziye et al., 2016). In general, sex differences in human physical dimensions become larger with onset of puberty (Matzdorff, 1967). Moreover, studies note considerable digit length fluctuations with male children's digit lengths surpassing females' only after the age of 12 years (Gillam et al., 2008; Manning & Fink, 2018). Overall, there is a lack of studies investigating digit and hand lengths during infancy, specifically with a rationale of investigating sex differences as most studies focus on anthropometric (see, e.g., Brandt et al., 1990; Marshall, 1975; Matzdorff, 1967) or forensic information (see, e.g., Aboul-Hagag et al., 2011; Kanchan & Krishan, 2011; Krishan et al., 2011).

Regarding digit ratios, the hypothesized sex difference in 2D:4D (i.e., males < females) was small and only marginally significant for right 2D:4D (p = .061, d = 0.14) and averaged 2D:4D (p = .092, d = 0.12), while it was non-significant for left 2D:4D (p = .329, d = 0.07). This aligns with a broad body of literature concentrating on 2D:4D during infancy and early childhood (Alexander & Saenz, 2012; Barrett et al., 2020; Hwang et al., 2014; Knickmeyer et al., 2011; Lutchmaya et al., 2004). In fact, in the current study, digit ratios that comprised 5D as a denominator resulted in the largest effect sizes (all p <.001, d = 0.25 to 0.40), which aligns with Kumar et al. (2017), although, the authors measured digit lengths

dorsally, which crucially differs from the commonly cited technique using the ventral surface of the hand (see, e.g., Kemper & Schwerdtfeger, 2009). Other studies also found significant sex differences in other digit ratios (Manning, 2012; McIntyre et al., 2005), but did not mention the usefulness of digit ratios comprising 5D. Instead, both argued that 3D:4D may be more applicable in younger cohorts (Manning, 2012; McIntyre et al., 2005), a statement that cannot be reinforced by the results of the current study (all $p \leq .060$, $d = 0.14$ to 0.21). McIntyre et al. (2005) argued that the measurement of longer digit segments accounts for less measurement error, which should be investigated in future studies.

Loehlin et al. (2009) introduced relative digit lengths in an adult sample and found the largest effect size in rel2, while rel4 showed only small sex differences. Further, males had larger rel2 and rel3, whereas females had larger rel4 and rel5 (Loehlin et al., 2009). In the current study, we could not replicate this finding, as male children had larger rel5 (all $p < .001$, $d = 0.32$ to 0.43), while female children had larger rel2 (all $p \leq .020$, $d = 0.17$ to 0.20) and rel3 (all $p < .001$, $d = 0.28$ to 0.33). Further, sex differences were not evident for rel4 (all $p \geq .266$, $d = 0.02$ to 0.08).

Lastly, no sex differenced were found in $D_{[r\text{-}l]}$ in the current sample ($p = .379$, $d = 0.06$), contradictory to Manning et al. (2019) who found sex differences in an adult sample with self-measured, rather than expert-measured, digit lengths. Other studies that employed expert-measured digit lengths similarly did not report significant sex differences in $D_{[r\text{-}l]}$ (Coyne et al., 2007; Richards et al., 2019).

Conclusion

The results indicate that a sexual dimorphic pattern in hands and digits is apparent as early as 6 months of age. Surprisingly, the sex difference in infant's 2D:4D was not significant, however, is commonly cited as evidence for its association with prenatal T. Accordingly, the current results question the applicability of 2D:4D in younger cohorts. Hand and digit measures of the current study were highly comparable in terms of the measurement technique, which was one major problem in the available literature. Lastly, relevant methodological implications and a valuable reliability analysis promote the usefulness of indirect scans and highlight relevant factors (e.g., standardized measurement protocols) that should be considered when obtaining indirect measures in very young cohorts.

Study 3

The association between 2D:4D digit ratio and sex-typed play in children with and without siblings

(Ernsten et al., 2024, Scientific Reports)

Study 1 could not support the underlying hypothesis of 2D:4D to be associated with prenatal T and Study 2 has found that 2D:4D was only able to a small or nonsignificant extent to differentiate between girls and boys of 6 months while other digit ratios revealed larger effect sizes. Nevertheless, 2D:4D is still the commonly used marker for prenatal T exposure (Manning & Fink, 2023). To establish some sort of criterion validity, research also focused on establishing associations between 2D:4D and human behavior that is also known to differ between males and females and presumably are linked to the effect of prenatal androgens like T. In children, sex typed play is often investigated as it reveals large sex differences up to Cohen's *d* = 1.60 (Davis & Hines, 2020) and, as also non-human primates exhibit sex-typed play (Lonsdorf, 2017; Marley et al., 2022), a biologic component seems rationale. Further, clinical samples with altered prenatal T concentrations exhibit sex-typed play behavior that is in accordance with their hormonal profile during gestation: individuals with CAH, a clinical condition that results in significantly heightened T levels during gestation, exhibit more male-typed play behavior compared to healthy controls (Cohen-Bendahan et al., 2005; Hines, 2003), while individuals with CAIS, a clinical condition that is associated with a complete insensitivity of the androgen receptor gene, exhibit more female-typed play behavior (Cohen-Bendahan et al., 2005). Some studies report evidence for an association between 2D:4D and sex-typed play in healthy pre-school children, with smaller 2D:4D related to more male-typical play behavior, however, only in boys (Hönekopp & Thierfelder, 2009; Mitsui et al., 2016), only in girls (Körner et al., 2020) or in both sexes (Wong & Hines, 2016). Other studies found no significant association in either sex (Barrett et al., 2020). Further, in humans, the development of sex-typed behavior such as sex-typed play is assumingly not only influenced by biologic factors. Bussey and Bandura (1999) proposed a social cognitive theory and note, amongst psychological

factors of the individuum, the importance of social influences or social models on shaping sex-typed behavior. Siblings potentially serve as such models for social learning (Bandura, 1971; Golombok & Hines, 2002). Only a few of the aforementioned studies included siblings as a factor in their analysis and found a significant influence of older siblings on sex-congruent play behavior (Körner et al., 2020; Mitsui et al., 2016), whereas others did not find any effect of siblings on sex-typed behavior (Hönekopp & Thierfelder, 2009). Further, none of the studies has investigated the association between 2D:4D and sex-typed play in singleton children who are potentially less influenced by social learning. This was shown by Rust et al. (2000) who found sex-congruent effects of siblings on gender-typed behavior while singletons exhibited less sex-congruent behavior.

The aim of study 3 was to investigate the association of children's 2D:4D during infancy and their subsequent sex-typed play 3.5 years later with an additional focus on the influence of siblings.

Methods

For study 3, a subset of the sample that also participated in study 2 was used for data analysis. As only datasets with valid hand scans for both the right and left hands, along with corresponding parental ratings of children's sex-typed play, were utilized, data from 505 children (240 female, 265 male) were included in the analysis. Hand scans of infants were obtained using a flatbed scanner when children were 6 months of age. All scans were rated twice by independent raters blind to the sex of the infants using the computer program *Autometric* (DeBruine, 2004). Intra-class correlations between both ratings were high (.95-.97) and accordingly both ratings of digit lengths were averaged to compute the 2D:4D digit ratio. Parents of infants were contacted 3.5 years later to participate in an online questionnaire that consisted of the Pre-School Activities Inventory (PSAI; Golombok & Rust, 1993), which assesses children's sex-typed play by parents' reports and self-designed questions regarding the number of older and younger sisters and brothers as well as the attendance of pre-school. Of all included children, 134 children (66 female, 68 male) were singletons and 98.6% of children attended preschool.

To test for sex differences, independent samples *t* tests were conducted with sex (female versus male) as dependent variable and right 2D:4D, left 2D:4D or PSAI score respectively as independent

variable. The association between (right and left) 2D:4D and PSAI was tested using one-tailed Pearson correlations separated by sex in the overall sample and in the subsample of singleton children. Lastly, a multiple regression model with the PSAI as criterion and sex, right 2D:4D, left 2D:4D, older sisters (dummy coded as 0 = no older sisters, 1 = at least one older sister), younger sisters (dummy coded as 0 = no younger sisters, 1 = at least one younger sister), older brothers (dummy coded as 0 = no older brothers, 1 = at least one older brother) and younger brothers (dummy coded as 0 = no younger brothers, 1 = at least one younger brother) was performed to test for the influence of siblings.

Results and discussion

It could be shown that boys had significantly smaller right 2D:4D (p = .048, d = 0.16), but not left 2D:4D (p = .180, d = 0.08) compared to girls and that boys and girls significantly differed regarding the PSAI score. Boys tended to play more masculine and girls more feminine as rated by their parents (p <.001, d = 2.60). No significant association between 2D:4D and sex-typed play could be found neither for boys nor for girls in the overall sample (all $p \leq .214$). The same correlations were performed in a subsample of singleton children. A small significant correlation could be found for boys right (r = -.25, p = .020) and left 2D:4D (r = -.26, p = .015) with the PSAI score. No significant correlation between 2D:4D and PSAI was found for girls (all $p \leq .135$). The multiple regression model was significant, F (7, 497) = 127.79, p <.001, R^2 = .64; and a significant influence on the PSAI could be found for sex (β = -.79, p <.001), younger sisters (β = -.06, p = .042) and older brothers (β = .08, p = .003). To disentangle the significant influence of sex, multiple regression models were performed separately for boys and girls using the same criterion and same predictors. In boys, the overall model was significant, F (6, 258) = 4.09, p <.001, R^2 = .09. Further, having sisters, both younger (β = -.16, p = .012) and older (β = -.19, p = .004), promoted feminized play behavior while the existence of older brothers was associated with more male typical play behavior in boys (β = .16, p = .011). In girls, the overall model was not significant, F (6, 233) = 0.63, p = .705, R^2 = .02, and none of the predictors was significant (all $p \geq .093$).

In sum, this study underscores the finding of only small or non-significant sex differences in 2D:4D in very young children, which is in line with a broad body of literature (Barrett et al., 2020;

Knickmeyer et al., 2011; McIntyre et al., 2005) as well as with results of Study 2. Further, the assumption of sex-typed play during early childhood could also be promoted, which aligns with several other studies indicating that large sex differences exist in sex-typed play measured with the PSAI (Barrett et al., 2020; Davis & Hines, 2020; Hines, 2020b; Hönekopp & Thierfelder, 2009; Körner et al., 2020; Mitsui et al., 2016; Wong & Hines, 2016), as well as when measured with other instruments (Alexander & Saenz, 2012; Knickmeyer et al., 2005; Lamminmaki et al., 2012; Spencer et al., 2021; van de Beek et al., 2009). Further, contrary to studies that found a significant association between PSAI and 2D:4D in boys (Hönekopp & Thierfelder, 2009; Mitsui et al., 2016), girls (Körner et al., 2020) or both sexes (Wong & Hines, 2016), in the current study, no such association could be found in the overall sample. The finding of the current study aligns with Barrett et al. (2020), who also did not find any significant association between 2D:4D and sex-typed play. It is commonly assumed that sex-typed play relates to sex hormones and is influenced by prenatal androgen action. This could also be shown by studies that found a significant association between sex-typed play and prenatal T measured in amniotic fluid (Auyeung et al., 2009), prenatal T measured in maternal blood serum during gestation (Hines et al., 2002) or free urinary T in children after birth (Lamminmaki et al., 2012). In all studies, higher T related to more male-typed play behavior (Auyeung et al., 2009; Hines et al., 2002; Lamminmaki et al., 2012). However, there are also several studies that did not find any significant association between sex-typed play and amniotic T (Knickmeyer et al., 2005; Spencer et al., 2021; van de Beek et al., 2009) or free salivary T in children (Alexander & Saenz, 2012). Although, there is sufficient evidence that sex-typed play in children also originates from biological factors (Alexander & Wilcox, 2012; Lonsdorf, 2017; Marley et al., 2022; Todd et al., 2018), other factors may be important to consider. Social factors are known to affect children's behavior (Bandura, 1971; Golombok & Hines, 2002) and possibly allow an explanation for the heterogeneous findings of the studies noted above.

The association between 2D:4D and PSAI did not reach significance in the overall sample, but in the subsample of singletons, a significant association could be found in boys. Both correlations (i.e., right and left 2D:4D with PSAI) could be interpreted as small (Cohen, 1988), however, emphasize the

possible importance of social influences on sex-typed play. In the current study, these social factors (i.e., siblings) potentially outran the proposed biological effects on sex-typed play as reflected by 2D:4D. As children with siblings are confronted with social models, it is possible that they adapt the observed play behavior of their siblings or interact with sex-typical toys (Kuchirko et al., 2021; Rust et al., 2000) and therefore develop play behavior that is congruent to their sibling's sex. In absence of siblings, children possibly are less influenced by social models or toys and therefore exhibit less sex-congruent behavior, which could be shown by Rust et al. (2000). In the study of Rust et al. (2000), children with and without siblings were compared regarding their sex-typed play behavior. It could be shown that older same-sex siblings promoted sex-congruent behavior, while older other-sex siblings promoted sex-incongruent behavior (Rust et al., 2000). Singleton children exhibited more sex-typical behavior compared to children with other-sex siblings and less compared to children with same-sex siblings (Rust et al., 2000). These findings align with the findings of the current study, as the existence of older brothers was associated with more male-typed play behavior in boys, as well as a significant influence of older and younger sisters on boys' female-typed play behavior. It was quite surprising that also younger sisters had a significant influence of sex-typed play behavior, as the literature specifically emphasized the effect of older siblings (Rust et al., 2000). There is only few research focusing on the influence of younger siblings. Hughes et al. (2018) assumed that children with younger siblings eventually exhibit more caregiving behavior. Caregiving is a behavior that is often interpreted as more female (Esplen, 2009), which can also be seen in some items used in the PSAI (see, e.g., item 1 and 6 of the activities subscale of the PSAI "Playing house [e.g., cleaning, cooking]", "Pretending to be a family character [e.g., parent]" that are both summed up to the female score; Golombok & Rust, 1993) leading to a feminized PSAI score. Although, this finding did not apply for the influence of younger brothers in the current study, it could be an explanation for the "feminizing" effect of younger sisters on their brothers' sex-typed play behavior. The results of the current study indicated a larger effect of siblings while 2D:4D showed no association to children's sex-typed play behavior, except for singleton boys.

Conclusion

The current study emphasized the possible weakness of 2D:4D as a marker for prenatal T in very young cohorts as the sex differences was only small or non-significant. The sex difference in pre-school children's sex-typed play could be promoted with overall large effect sizes. The correlation between 2D:4D and PSAI was only evident in singleton boys, but not in the overall sample. This could highlight the importance of social influences of sex-typed play, here as indicated by siblings. Lastly, combining 2D:4D and siblings in a regression model additionally underlined the effect of siblings on children's sex-typed play, with larger effects in boys.

Study 4

Preliminary findings on the association between maternal salivary and hair cortisol and the mother-infant-interaction during the early postpartum period

(Ernsten et al., 2024, accepted for publication after minor revision, Psychoneuroendocrinology)

Cortisol as a stress hormone can lead to a change from sensitive to more habitual memory systems, which presumably leads to impairments in behavioral flexibility (Schwabe & Wolf, 2013). During mother-infant-interactions, mothers need to adequately perceive, interpret and react to infants' signals in a timely manner, which is also described as maternal sensitivity (Shin et al., 2008). Evidence indeed suggests a negative effect of cortisol on maternal sensitivity after 3 months postpartum, whereas during the first 3 months cortisol could be associated with higher alertness of mothers to infant cues (Almanza-Sepulveda et al., 2020). In early infancy, the infant relies on maternal sensitivity to cope with different stressors in a dyadic regulatory approach (DiCorcia & Tronick, 2011), which fosters infant's own self-regulatory behavior (Blair & Ku, 2022). Evidence suggests that experimentally stressed mothers exhibit more negative interaction patterns during a standardized laboratory observation with their 4 month old infants and that these infants were also more irritated, showed heightened negative affect and crying compared to infants of mothers in the non-stressed condition (Mueller et al., 2021). These authors, however, did not control for changes in cortisol levels in response to the stressor (Mueller et al., 2021), for example, by assessing commonly used saliva samples (Kirschbaum & Hellhammer, 2008). Tarullo et al. (2017) found that cumulated maternal cortisol levels measured in hair over a given period of time (1 cm of hair refers to approximately one month; Meyer & Novak, 2012), related to more negative and less positive mother-infant-interaction at 7 months during a free-play.

In sum, research suggests that cortisol potentially impairs the mother-infant-interaction, potentially due to less sensitive maternal engagement during the interaction (Barrett & Fleming, 2011). To date, there are only few studies that investigated the association between the mother-infant-interaction and cortisol using endocrine markers in standardized laboratory assessments. The aim of study 4 was to investigate HCC and SCC as markers of maternal long-term and situational cortisol levels and their relation to the mother-infant-interaction in a standardized laboratory observation, the face-to-face still-face paradigm (FFSF; Tronick et al., 1978). The FFSF is a useful tool for the investigation of dyadic regulation according to the mutual regulation model (DiCorcia & Tronick, 2011) as the maternal still-face (i.e., maintaining a neutral face and stopping all interactions with the infant) evokes a mismatching state in dyadic interaction (Tronick et al., 1978). After mothers resume the interaction with their infants, an investigation of dyadic regulatory behavior is possible. In study 4, it was hypothesized that maternal SCC and HCC can be associated with less positive and more negative maternal, infant and dyadic behavior during the last episode of the FFSF.

Methods

The study involved mother-infant-dyads when infants were 4 months of age. As part of a larger study investigating the effect of stress on the mother-infant-interaction and differences between mothers of preterm and full-term infants, only full-term infants and their mothers were included in the current data analysis. Out of 76 dyads, 59 full-term dyads met eligibility criteria for data analysis (exclusion criteria were deviations from the study protocol, technical issues, missing HCC or SCC and excessive infant crying, i.e., >80% during observation). Mother-infant-interactions were videotaped with one camera each oriented towards the mother and the infant for a 5-minute free-play followed by the FFSF (Tronick et al., 1978). The mother sat in a chair and the infant in a slightly elevated infant seat so that mothers and infants were at eye level. Mothers were instructed to interact with their infant in a normal manner for the first 5 minutes and then for the first 2 minutes (play episode) of the FFSF, followed by an acoustic signal that indicated a switch to the following phase (still-face episode), where mothers maintained a neutral face and stopped any interaction with their infant for the following 2

minutes. After an additional acoustic signal, the mother could resume the normal interaction for 2 minutes (reunion episode). For the current analysis, only the last episode of the FFSF was analyzed. Behavior was coded using the German version of the Infant and Caregiver Engagement Phases Revised (Reck et al., 2009), focusing on infant social positive engagement, infant social monitor, infant protesting behavior, infant self-comforting behavior (oral and self-clasp), maternal social positive engagement, maternal social monitor/ positive vocalizations and dyadic eye contact. The description of each code can be seen in Reck et al. (2009). Infant social positive engagement and infant social monitor were summed up to infant positive affect, reflecting a cumulated measure for infants' positive and mother-directed engagement. Maternal social positive engagement and maternal social monitor/ positive vocalization were analogously summed up to reflect maternal positive affect directed at the infant. Further, instances where mothers and infants simultaneously displayed positive affect were coded as dyadic match (further referred to as dyadic match). Lastly, the interval between the onset of the reunion episode and the first dyadic match was assessed reflecting the duration until the first matching state occurred after maternal unresponsiveness (further referred to as interactive reparation). Saliva samples collected prior to and after the FFSF to assess SCC were used as a baseline measure for maternal situational cortisol prior to the mother-infant-interaction. The area under the curve with respect to ground (AUC_G; Pruessner et al., 2003) was calculated which accounted for the time lag between the first and second saliva sample. HCC were measured from a 3 mm thick hair strand that was cut off near the mother's scalp (Meyer et al., 2014). The first 3 cm of this strand were used for analysis, reflecting cumulated cortisol levels over the past 3 months (Meyer & Novak, 2012). The mothers filled out additional questionnaires regarding sociodemographic information and answered standardized questionnaires, these were not included in the current analysis.

Infants' responses to maternal unresponsiveness during the FFSF, indicated by changes in infants' negative and positive affect as well as infants' self-comforting behavior between phases, were analyzed using repeated measures ANOVA. To investigate the influence of HCC and SCC, block-wise hierarchical multiple regression models were tested. Sociodemographic factors (i.e., maternal age,

parity, infant gestational age, infant sex) were included as control variables in a first step and HCC and SCC were entered as factors in a second step, assessing changes in explanation of variance in infant positive affect, protesting behavior, maternal positive affect, dyadic match, interactive reparation and dyadic gaze.

Results and discussion

Infants showed a decrease in social positive engagement to a moderate extent (p <.001, η^2 = .12) with significant differences between play and still-face, as well as play and reunion and no significant differences between still-face and reunion. Further, there was an overall large effect for infants' protesting behavior (p <.001, η^2 = .33), with infants significantly increasing their protesting behavior from play to still-face, play to reunion and still-face to reunion. Lastly, infants' self-comfort did not change significantly for oral-comfort (p = .728, η^2 = .01), however, there was a significant difference for self-clasp behavior (p <.001, η^2 = .19), with a significant increase from play to still-face and a significant decrease from still-face to reunion. Therefore, the overall robustness of the effect of the FFSF could be confirmed.

The focus of the study was on the influence of cortisol reflected by two endocrine markers HCC and SCC on different maternal, infant and dyadic behaviors in the standardized FFSF. The addition of HCC and SCC as predictors led to a significant increase in explanation of variance for infant positive affect (ΔR^2 = .15, p = .007), infant protesting behavior (ΔR^2 = .15, p = .011), dyadic match (ΔR^2 = .16, p = .005) and dyadic eye contact (ΔR^2 = .11, p = .039), which could be interpreted as low to moderate. Maternal HCC was no significant predictor for none of the observed behaviors (all $p \geq .205$). Higher SCC significantly predicted less infant positive affect (β = -.37, p = .004), higher infant protesting behavior (β = .36, p = .008), less dyadic match (β = -.39, p = .003) and less dyadic eye contact (β = -.33, p = .014). Maternal positive affect and interactive reparation were unaffected by any of the predictors. Further, infant gestational age at birth and infant sex assigned birth seemed to be relevant influencing factors for the significant regression models as they also contributed to the explanation of variance. Infant gestational age at birth was significantly related to infant positive affect (β = -.41, p = .001) and dyadic

match (β = -.31, *p* = .015). Infant sex assigned at birth was related to infant protesting behavior (β = .29, *p* = .031) and dyadic match (β = -.26, *p* = .038). Maternal age and parity had no significant influence on the behavior of interest.

The study suggests that elevated cortisol levels may disrupt mother-infant interaction, leading to increased infant negativity and reduced infant positivity. This could also be shown by other studies that used an experimental maternal stressor prior to the mother-infant-interaction, although, these authors did not measure cortisol levels (Mueller et al., 2021). It has been discussed that cortisol negatively impacts behavioral flexibility (Schwabe & Wolf, 2013), but it has not been examined yet if this also applies to maternal abilities to engage in sensitive caregiving. Future studies could integrate cortisol measurements and standardized experimental stress manipulations to investigate more causal effects of cortisol levels on the mother-infant-interaction as the current results allow only preliminary and correlational interpretation. Another interpretation of the current findings may be that mothers with higher SCC were stressed due to the arrival to the laboratory and subsequent physical activation (Hellhammer et al., 2009; Kirschbaum & Hellhammer, 2008). Further, mothers with higher SCC eventually exhibit less optimal interaction with their infants in a more general way. It could be shown that elevated maternal cortisol levels during pregnancy can be associated with more irritable infants after birth (Takegata et al., 2021) and a reciprocity between elevated cortisol levels and infant irritability cannot be ruled out. This may be underlined by the null effects of SCC on maternal positive affect, which is contrary to other studies reporting negative effects of elevated cortisol levels on maternal sensitivity (Finegood et al., 2016) or intrusiveness (Mills-Koonce et al., 2009). The null-effects on maternal positivity can also be interpreted in light of a possible lack of environmental validity due to the laboratory assessment, which may led to overall higher positivity in maternal interactive behavior (Zegiob et al., 1975) or socially desirable maternal behavior (Belsky, 1980). Finally, elevated cortisol levels have also been reported to relate to maternal attention towards the infant and the infant's signals (Fleming et al., 1997), which could align with the overall positive maternal engagement during reunion in the current study.

No significant associations could be found for maternal HCC, which contradicts other studies that found negative effects of elevated HCC on mother-infant-interaction (Tarullo et al., 2017). Maternal mental health seems to be a relevant factor that should be incorporated when investigating HCC, as Nystrom-Hansen et al. (2019) found that elevated HCC related only to maternal intrusiveness in mothers with a mental illness that was evident during the third trimester as well as 4 months postpartum (Nystrom-Hansen et al., 2019). Khoury et al. (2020) found comparable study results between elevated HCC that related only to maternal withdrawal when mothers showed depressive symptoms during the third trimester or 4 months postpartum, whereas maternal intrusiveness related only to maternal HCC when depressive symptoms were low. Otherwise, chronic mental illnesses can also downregulate HPA axis activity resulting in overall lower cortisol levels (Pochigaeva et al., 2017). The HCC of mothers in the current sample were relatively low compared to other studies (Horan et al., 2022; Kirschbaum et al., 2009), which may reflect such long-term downregulation of HPA axis activity. But, as the current sample was relatively high in socioeconomic status and educational levels and lived mostly in partnerships, it is more likely that mothers in the current study experienced overall only low levels of stress, which was reflected in lower HCC. For future studies, other factors like maternal mental health and other risk factors that potentially lead to a chronic up- or down-regulation of the HPA axis should be considered.

Additional factors have been incorporated in the current analysis, as evidence suggested possible influences of maternal age (Bornstein & Putnick, 2007; Bornstein et al., 2019; Bornstein et al., 2006), parity (Fish & Stifter, 1993; Hsu & Fogel, 2003), infant gestational age at birth (mostly in the context of prematurity, see, e.g., Bilgin & Wolke, 2015; Korja et al., 2012) and infant sex assigned at birth (Barbosa et al., 2019; Weinberg et al., 1999) on the mother-infant-interaction. The current results indicated that higher infant gestational age as well as male infant sex assigned at birth significantly related to less dyadic interaction and positive infant affect, as well as more negative infant affect during mother-infant-interaction. Maternal factors like maternal age or parity revealed no association to any of the observed behaviors. Research on the importance of infant gestational age at birth in full-term children is scarce, whereas male sex could be associated with higher infant irritability in dyadic

interaction in other studies (Weinberg et al., 1999). It could be beneficial to include both factors a source of variance in the parent-infant-interaction in future studies.

Conclusion

The current study gives valuable insights into the hypothesized importance of maternal cortisol levels and the mother-infant-interaction. Conclusive literature is currently lacking, specifically incorporating endocrine markers for the assessment of cortisol in a situational and long-term context. Although, the current results are only small in effect sizes and rather reflect correlational results that need to be addressed in a more standardized context, it could be shown that especially situational cortisol levels potentially impair the mother-infant-interaction. For the investigation of long-term, cumulated stress, other important factors like mental illnesses should be considered in future studies.

General Discussion and Conclusion

The aim of the current study was to evaluate the relevance of two key hormones in early human development. The first part investigated 2D:4D as a marker for prenatal T in three studies. The second part aimed to investigate the association between maternal momentary and long-term cortisol levels on mother-infant-interaction in a standardized FFSF observation.

The first part of the current work presented three studies that investigated the applicability and usefulness of 2D:4D as a marker for prenatal T. There is much research on the association between 2D:4D and several characteristics that are assumed to be related to the influence of prenatal T, however, studies investigating the theoretical foundation by obtaining more direct information about prenatal T levels in humans and the association with 2D:4D are scarce. In Study 1, rare amniocentesis data could be used to investigate the more direct association between amniotic T and children's 2D:4D at 5, 9, 20, 40 and 70 months. No significant association could be found, a finding that was limited to the male subsample. A replication with male and female children's amniotic T should be aimed for in future studies. Other studies that also used amniotic T reported conflicting results (Lutchmaya et al., 2004; Richards et al., 2020a; Ventura et al., 2013), leaving behind significant uncertainty and heterogeneity concerning the validity of 2D:4D. The sex difference in 2D:4D was evident for all measurement time points which was surprising as the vast amount of studies investigating 2D:4D in very young cohorts reported only small or non-existent sex differences (Knickmeyer et al., 2011; McIntyre et al., 2006; McIntyre et al., 2005; Trivers et al., 2006; Wong & Hines, 2016). Further, a significant age effect contradicts other studies that described 2D:4D to be independent of age (Manning & Fink, 2018), while it is in accordance with a recent study suggesting stable sex differences and age effects in 2D:4D between pre-school age and adolescence (Eichler et al., 2023). The more direct approach using amniotic fluid data can be considered as an important contribution of Study 1 to the existing literature as this approach is currently lacking in research on 2D:4D, yet, being limited to the male subsample. The results of Study 1 weaken the theoretical foundation of 2D:4D as a hypothesized reflection of prenatal T levels, although, consistent sex differences in 2D:4D could be found.

In Study 2, a large data set of 6 months old children was used to investigate different hand and digit measures, as the available literature in young cohorts is scarce and differs in terms of measurement methods. This makes the comparability of different studies rather difficult. In Study 2, 2D:4D was one of the measures yielding smallest and non-significant sex differences. While this finding should be also interpreted in light of possible methodological issues that can arise when obtaining indirect hand scans in very young children (e.g., shifting of landmarks due to higher amounts of body fat, compliance during measurements), other digit ratios indeed detected moderate sex differences. Other authors like McIntyre et al. (2005, 2006) already proposed that it may be more appropriate to use 3D:4D in younger cohorts, as longer digit segments possibly reduce measurement errors. However, it should be noted that these authors used radiographs instead of indirect measurement (McIntyre et al., 2006; McIntyre et al., 2005), which impairs comparability. In Study 2, effect sizes observed for 3D:4D were larger compared to those for 2D:4D, however, effects were still small. Notably, the most pronounced (i.e., moderate) effects were evident in digit ratios with 5D as the denominator, contrasting the findings of Manning (2012), who did not find significant sex differences in digit ratios comprising 5D (derived from direct measurements using a caliper). While the reasons for the lack of significant sex differences in 2D:4D remain speculative, Study 2 provides valuable insights into different hand and digit measures that were able to detect sex differences in young children.

Lastly, Study 3 aimed to investigate the applicability of children's 2D:4D and its association with sex-typed behavior, which is proposed to relate to prenatal T. Both, the sex difference in 2D:4D and the correlation between 2D:4D and sex-typed play behavior did not reach statistical significance. This directly contradicts other studies that found 2D:4D to be related to sex-typed play in children (Hönekopp & Thierfelder, 2009; Körner et al., 2020; Mitsui et al., 2016; Wong & Hines, 2016). Only in a subsample of singleton boys, the association was significant. The influence of siblings was a significant factor affecting children's sex-typed play and it could be found, in accordance with other studies (Rust et al., 2000), that children showed sex-typed play behavior that was congruent to the sex of their siblings. However, siblings possibly do not only affect sex-typed play behavior, as studies suggest that having,

especially, more older brothers decreases 2D:4D in children while the effect of older sisters remains contradictory (Kralik et al., 2019; Saino et al., 2006; Williams et al., 2000). Therefore, it is possible that the association between 2D:4D and sex-typed play in the overall sample was masked by the influence of birth order on 2D:4D. In fact, studies suggest 2D:4D is possibly affected more generally by hereditary or genetic factors that potentially outrun effects due to prenatal T or androgen sensitivity (Andrew et al., 2006; Loehlin & Medland, 2008; Richards et al., 2017; Voracek & Dressler, 2007, 2009).

Some limitations and methodological considerations should be noted that affect the results of Studies 1, 2 and 3. Some authors emphasize that while investigating sex differences in 2D:4D or 2D:4D-behavior correlations, several other factors have to be taken into account that have been considered as potential sources of variance in 2D:4D: ethnicity (Butovskaya et al., 2021; Manning et al., 2007), hand preference (Gillam et al., 2008, but see Lupu et al., 2023) or overall body size (Lolli et al., 2017). Further, other factors can affect T levels during pregnancy, like maternal age (Carlsen et al., 2003; Kallak et al., 2017). None of the current studies did account for these factors. Further, sex-typed behavior is known to be also affected by psychological factors, as preschool-aged children already have rudimentary cognitive theories about gender roles (Golombok & Fivush, 1994; Golombok & Hines, 2002). Therefore, a bio-psycho-social model may be more appropriate, whereas Study 3 only considered biological and social factors.

Nevertheless, summarizing the findings of Studies 1, 2 and 3 as well as considering existing evidence that has been discussed above, evidence leads to the necessity for a critical re-evaluation of 2D:4D as a candid marker for prenatal T exposure and being used as a surrogate. Animal studies indicate that manipulation of prenatal T levels can also affect limb development as well as 2D:4D (Suchonova et al., 2019; Talarovičová et al., 2009; Zheng & Cohn, 2011), but evidence in humans remains significantly more uncertain (Manning & Fink, 2023; Sorokowski & Kowal, 2024). This uncertainty also arises from the results of Studies 1, 2 and 3, especially regarding sex differences (Study 1 found sex differences in 2D:4D, while Study 2 found it to some extent and Study 3 did not find a sex difference) or the association between 2D:4D and sex-typed play in children (in Study 3, an association could be found for singleton

boys, however, not in the overall sample). The current work stresses the need for meaningful research that aims to evaluate the true association between prenatal T levels and 2D:4D, especially, as differences in male and female fetuses' prenatal T levels are large in effect size (Abramovich, 1974; Auyeung et al., 2013; Robinson et al., 1977) while sex differences in 2D:4D remain small to moderate (Hönekopp & Watson, 2010; Richards et al., 2020b). Hönekopp and Watson (2010) already noted that there seems to be another variable that potentially impacts the relationship between 2D:4D and prenatal T and that future research should address this specific circumstance. However, since the publication of Hönekopp and Watson (2010) many studies used 2D:4D as a marker for prenatal T and only few studies focused on more basic research that aims to validate 2D:4D. A detection of other sources that explain variance in 2D:4D are highly needed and would make assumptions concerning associations between 2D:4D as a reflection of prenatal T and other human characteristics much more reliable.

The second part of the current work focused on maternal cortisol and its association with mother-infant-interaction. Study 4 used a standardized face-to-face still-face procedure (Tronick et al., 1978) and found that maternal salivary cortisol levels related to less positive infant affect and less matched behavior between mothers and infants and further with more infant protest behavior. Maternal hair cortisol did not show any association to infant, maternal or dyadic behavior. Infant gestational age at birth and sex assigned at birth were relevant confounding factors for the mother-infant-interaction. The negative impact of maternal salivary cortisol on mother-infant-interaction can be derived from other studies (Finegood et al., 2016; Thompson & Trevathan, 2008), yet, studies focusing on causal effects are still lacking. Therefore, the current results must be interpreted as preliminary and more research using standardized study designs are highly needed. Surprisingly, maternal hair cortisol did not relate to any mother-infant-behavior in Study 4, conflicting with other studies that reported evidence for negative effects of cumulated maternal cortisol (Khoury et al., 2020; Nystrom-Hansen et al., 2019; Tarullo et al., 2017). These findings were mostly associated with maternal mental health issues, like depression (Khoury et al., 2020; Nystrom-Hansen et al., 2019), which seem to

be relevant factors to incorporate when investigating maternal hair cortisol. While it would be non-permissible to conclude impairments in maternal caregiving solely based on momentarily elevated cortisol levels, Study 4 allows to infer directions for future research, for example, the necessity to combine cortisol measurements and standardized stressor paradigms as used by Mueller et al. (2021) and Tronick et al. (2021). Following this, it would be more appropriate to draw conclusions about causal effects of maternal cortisol elevation on impaired mother-infant-interaction.

Outlook

Future directions for further research can be derived from the results and limitations presented in the current work. In order to stick to the structure of the work, in a first part, future directions for 2D:4D research will be presented. The second part then focuses on research on maternal cortisol in the context of mother-infant-interaction. Lastly, an outlook will be given on how to combine both topics and why it can be helpful to consider both hormones, T and cortisol, in related research.

Future directions of 2D:4D research

The main aim of future 2D:4D research should be to conduct more grounded research focusing on the hypothesized relation between prenatal T and 2D:4D. As Hönekopp and Watson (2010) already noted in their work, it may be crucial to detect the source of variance in 2D:4D that cannot be explained by prenatal T so that future studies can control for this influence. Further, caution should be applied in terms of age. Although, some authors claim 2D:4D to be independent of age (Manning & Fink, 2018), the vast amount of studies investigating 2D:4D in different age groups indicates that, specifically in younger cohorts, 2D:4D varies by age and tends to become larger with age (Knickmeyer et al., 2011; Manning, 2012; McIntyre et al., 2006; McIntyre et al., 2005; Trivers et al., 2006). Larger sex differences in adult cohorts could be the result of further sexual differentiation, possibly as pubertal hormones affect different human tissues and lead to a further differentiation in a sex-specific manner (Rogol et al., 2002). Králík et al. (2017) critically investigated changes in 2D:4D during puberty and found intraindividual variability to be a significant source of variation in 2D:4D: They further note the importance to consider sex-dependent differences in development (i.e., puberty in females often begins much earlier and females reach their final body composition in terms of height also earlier compared to males; Králík et al., 2017). As also noted in a prior section, many other factors should be considered when investigating 2D:4D, for example, birth order (Kralik et al., 2019; Saino et al., 2006; Williams et al., 2000), ethnicity (Butovskaya et al., 2021; Manning et al., 2007), hand preference (Gillam et al., 2008, but see Lupu et al., 2023) or overall body size (Lolli et al., 2017). Nevertheless, some authors still describe 2D:4D as a “biomarker” (Fonseca et al., 2022; Manning et al., 2014; Manning & Fink, 2023). Following

the definition of Califf (2018), a biomarker requires the properties to explain changes in a specific outcome measure and that the correlation between the biomarker and the outcome itself does not serve as evidence for the applicability of the biomarker. Considering 2D:4D, there is some evidence in animal research that manipulation of prenatal T concentrations (Suchonova et al., 2019; Talarovičová et al., 2009) or the androgen receptor leads to changes in 2D:4D (Huber et al., 2017; Swift-Gallant et al., 2021; Zheng & Cohn, 2011). However, these results pertain to rodents, not humans. Although, human 2D:4D seems to relate to the influence of prenatal or, more precise, amniotic T in some way (Sorokowski & Kowal, 2024), the current evidence seems insufficient to vindicate its properties as a biomarker.

Lastly, it should be noted that the endeavor to rely on 2D:4D in terms of investigating influences of prenatal T on human behavior or even clinical outcomes that can be associated with prenatal T, is highly reasonable. Ethical considerations do not allow to measure prenatal T in humans more directly during pregnancy and amniocenteses may be more and more replaced by first-trimester screenings (Nicolaides, 2005). This could lead to the sample of such amniocentesis data becoming increasingly more selected (for example, fetuses with suspected chromosomal abnormalities), which then impairs generalizability of data. 2D:4D may serve as a correlate of prenatal T, however, considerations that have been noted above need to be taken into account in future research.

Future directions on maternal cortisol in the context of mother-infant-interactions

One of the most important future directions for research on effects of cortisol during mother-infant-interactions should be the implementation of standardized stress manipulations for mothers prior to observing the FFSF. For example, Mueller et al. (2021) and Tronick et al. (2021) developed a stress paradigm which used infant crying noises to evoke a stress response in mothers, although, it has yet to be proven if this manipulation leads to an HPA axis activation. Alternatively, validated paradigms like the Trier Social Stress Test (Kirschbaum et al., 2008) could be used, as it reliably evokes a cortisol response in laboratory settings (Allen et al., 2017). This could allow to detect more causal effects of elevated cortisol levels on the mother-infant-interaction. For this, measurement protocols should be used to control for factors possibly influencing cortisol output, like day time, physical activity, smoking,

eating and drinking (Kirschbaum & Hellhammer, 2008). Further, infant characteristics should be equally incorporated as evidence suggests reciprocity between maternal stress, which potentially relates to maternal cortisol output, and infant characteristics like temperament, irritability and crying (Adler et al., 2017; Takegata et al., 2021). To investigate a possible reciprocity between long-term and acute HPA axis activation, HCC should also be assessed in future research. Correlations between hair and salivary cortisol measures are known to be moderate to large (D'Anna-Hernandez et al., 2011; Short et al., 2016; Singh Solorzano et al., 2023; Zhang et al., 2018). However, it is still not fully understood how long-term stress affects HPA axis activation and concurrent cortisol output (Miller et al., 2007; Rohleder, 2019; Young et al., 2020), which should be taken in mind.

Following a more systemic approach, it could be conclusive to additionally consider other caregivers, like partners (as well as other parents outside of traditional family constellation ideas), when investigating the interaction with an infant. Fathers have been found to significantly affect child socio-emotional development (see, e.g., Byrd-Craven et al., 2012; Roby et al., 2021; Rominov et al., 2016). Also maternal perceived stress has been found to be negatively associated with perceived support of the partner as well as with overall satisfaction with the partner (deMontigny et al., 2020), which could affect maternal cortisol levels and HPA axis activity. Adler et al. (1991) even found that lower perceived social support of mothers during pregnancy predicted mothers' perception of the child as more difficult at one year old. Lastly, Saxbe et al. (2014) found that HPA activity was interrelated in triadic pairs of mothers, fathers and their adolescent children, which should be taken into account while investigating the association between cortisol and the quality of family interactions.

Finally, it could be beneficial to gather information about maternal mental health and HPA axis activity in the context of stress already during pregnancy. It could be shown that maternal heightened cortisol levels during pregnancy, specifically during the last trimester, was associated with preterm pregnancies (Gilles et al., 2018), low infant birth weight (Cherak et al., 2018) and developmental problems of infants (Zijlmans et al., 2015). However, it should be noted that Zijlmans et al. (2015) reported in their systematic review that most studies did not find significant results. Studies that found

significant associations mainly found elevated maternal cortisol during pregnancy related to worse outcomes (Zijlmans et al., 2015). Further, studies suggest that infants of mothers with prenatally elevated cortisol levels are higher in irritability and display more negative affect in mother-infant-interactions (during a bathing situation, see de Weerth et al., 2003) or by mothers' reports (Nolvi et al., 2016). Maternal HPA axis activity and elevated cortisol levels during pregnancy have also been proposed to permanently alter the fetal HPA axis functioning (in the context of 'fetal programming', see, e.g., Galbally et al., 2021; Lewis et al., 2015). These are important findings that should be taken in mind as they potentially affect the overall mother-infant-interaction as well as maternal and infant HPA axis activity. Recruiting women in the last trimester of pregnancy, assessing their stress profile during pregnancy and investigating the early mother-infant-interaction during the postpartum would be beneficial to shed light on the complex association between maternal, fetal and subsequently infant HPA axis activity and the mother-infant-interaction.

Combining research on T and cortisol in the context of early development

To begin with a more general perspective, T in the context of sexual differentiation, sex differences in behavior and other sex-typed characteristics may be also important for parent-child-interactions. The literature reports that sex differences in socio-emotional behavior occur quite early in human development (McClure, 2000; Weinberg et al., 1999). Evidence suggests that these differences also relate to different parental interaction patterns with boys and girls (although, current research did not find robust evidence for sex-dependent behavior of parents, see Endendijk et al., 2016) and a biological origin for these sex differences has been discussed in the literature (Alexander & Wilcox, 2012). These sex differences were also apparent in Study 4, as infant's sex assigned at birth was a relevant factor for some aspects of the mother-infant-interaction, in a way that infant boys and girls differed in their affect and dyadic behavior after maternal unresponsiveness. Finally, sex differences have also been discussed for HPA axis function and its activity in response to psychosocial stress, however, with heterogeneous findings (Kudielka & Kirschbaum, 2005). These differences in HPA axis functioning may be shaped already in-utero, as Carpenter et al. (2017) summarized findings on the

effect of prenatal stress (i.e., increased exposure to maternal glucocorticoids) on programming of the fetal HPA axis and found that female fetuses seem to be more vulnerable to long-lasting alterations of HPA axis function compared to male fetuses.

Besides sex differences that may stem from the influence of prenatal T, fluctuating T in adults may be also an important factor for parents, as some evidence suggests that after birth T declines in fathers (Gettler et al., 2011), that T is lower in fathers compared to non-fathers (Gettler et al., 2018; Grebe et al., 2019) and that a T decline in mothers and fathers can be associated with more caregiving behavior, specifically in fathers (Kuo et al., 2018; Saxbe et al., 2017). These evidences are quite heterogeneous in terms of their applicability to mothers and fathers (see, e.g., Bos et al., 2018) and the direction of the effect (see, e.g., Bos et al., 2021; Cho et al., 2016). Other authors claim that effects of T on caregiving and parental behavior cannot be interpreted without incorporating interactions between T and cortisol (Beijers et al., 2022; Bos et al., 2018; Voorthuis et al., 2019); a claim that has been specifically investigated in the context of the dual hormone hypothesis.

The dual hormone hypothesis originated from the challenge hypothesis that has been originally tested for birds and explained season-dependent aggression versus breeding behavior of male birds due to changes in T levels (Archer, 2006). To apply this hypothesis to humans, it was necessary to adapt the hypothesis as T alone was not able to predict human social behavior in the context of dominance and reproduction and subsequently the dual hormone hypothesis was proposed (Knight et al., 2020; Mehta & Prasad, 2015). The dual hormone hypothesis addresses the interrelation of T and cortisol, as it describes that T is positively related to behaviors such as dominance or social aggression when cortisol levels are low, while the effect of T should be inhibited when cortisol levels are high (Mehta & Josephs, 2010; Mehta & Prasad, 2015). To date, many studies have been conducted to test the dual hormone hypothesis and methodological concerns have been raised (Grebe et al., 2019; Knight et al., 2020). In the context of caregiving behavior, some studies found that it may be beneficial to investigate the interrelation of T and cortisol and their findings aligned with the assumptions of the dual hormone hypothesis (Beijers et al., 2022; Bos et al., 2018; Voorthuis et al., 2019),

Outlook on own ideas for implementing future research directions

Some of the aspects noted above were already incorporated in a study of our research group to contribute to the proposed future directions. To gain a broader view of the mother-infant-interaction and how T and cortisol may be associated to dyadic and caregiving behavior, the study design of Study 4 was adapted, approved by the local ethics committee (study ID 1329) and is currently ongoing. Mothers give saliva samples for the investigation of momentary cortisol levels as well as saliva samples for the assessment of total T levels using liquid chromatography tandem mass-spectrometry (following the recommendations of Knight et al., 2020). Further, hand scans are obtained from both mothers and infants to obtain the 2D:4D digit ratio as a proposed marker for prenatal T exposure. To meet confounding factors that possibly impede the applicability of 2D:4D, several factors (e.g., birth order) will be incorporated in the final data analysis. Sex differences will be investigated for several factors of interest, such as maternal, infant and dyadic behavior during the FFSF as well as cortisol or T levels. Lastly, the role of partners will be addressed, as mothers receive a questionnaire on partner satisfaction. At the same time, partners of participating mothers are offered to take part in an online questionnaire that includes standardized measurements on partner mental health and bonding with the infant. Additional observation of partner-infant-interaction using the FFSF is also planned for future studies of our research group.

References

Aboul-Hagag, K. E., Mohamed, S. A., Hilal, M. A., & Mohamed, E. A. (2011). Determination of sex from hand dimensions and index/ring finger length ratio in Upper Egyptians. *Egypt J Forensic Sci*, *1*(2), 80-86. https://doi.org/10.1016/j.ejfs.2011.03.001

Abramovich, D. R. (1974). Human sexual differentiation--in utero influences. *J Obstet Gynaecol Br Commonw*, *81*(6), 448-453. https://doi.org/10.1111/j.1471-0528.1974.tb00494.x

Adler, N. E., Alkon, A., Bush, N. R., Caron, Z., Coccia, M., Coleman-Phox, K., Epel, E. S., Jones-Mason, K., Laraia, B. A., Thomas, M., & Wadhwa, P. D. (2017). Effects of pre- and postnatal maternal stress on infant temperament and autonomic nervous system reactivity and regulation in a diverse, low-income population. *Development and psychopathology*, *29*(5), 1553-1571. https://doi.org/10.1017/S0954579417001237

Adler, R., Hayes, M., Nolan, M., Lewin, T., & Raphael, B. (1991). Antenatal prediction of mother-infant difficulties. *Child Abuse & Neglect*, *15*(4), 351-361. https://doi.org/https://doi.org/10.1016/0145-2134(91)90020-E

Agorastos, A., & Chrousos, G. P. (2022). The neuroendocrinology of stress: the stress-related continuum of chronic disease development. *Molecular Psychiatry*, *27*(1), 502-513. https://doi.org/10.1038/s41380-021-01224-9

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Lawrence Erlbaum.

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

Alexander, G. M., & Wilcox, T. (2012). Sex Differences in Early Infancy. *Child Development Perspectives*, *6*(4), 400-406. https://doi.org/https://doi.org/10.1111/j.1750-8606.2012.00247.x

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The Trier Social Stress Test: Principles and practice. *Neurobiology of Stress*, *6*, 113-126. https://doi.org/https://doi.org/10.1016/j.ynstr.2016.11.001

Almanza-Sepulveda, M. L., Fleming, A. S., & Jonas, W. (2020). Mothering revisited: A role for cortisol? *Horm Behav*, *121*, 104679. https://doi.org/10.1016/j.yhbeh.2020.104679

Andrew, T., Cherkas, L. F., Kato, B. S., Paul, S. N., & Spector, T. D. (2006). Heritability of the Second to Fourth Digit Ratio (2d:4d): A Twin Study. *Twin Research and Human Genetics*, *9*(2), 215-219. https://doi.org/10.1375/twin.9.2.215

Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience & Biobehavioral Reviews*, *30*(3), 319-345. https://doi.org/https://doi.org/10.1016/j.neubiorev.2004.12.007

Arnold, A. P. (2009). The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Hormones and Behavior*, *55*(5), 570-578. https://doi.org/https://doi.org/10.1016/j.yhbeh.2009.03.011

Arnold, A. P., & McCarthy, M. M. (2016). Sexual Differentiation of the Brain and Behavior: A Primer. In N. D. V. D. W. Pfaff (Ed.), *Neuroscience in the 21st Century* (pp. 2139-2168). Springer Science + Business Media. https://doi.org/10.1007/978-1-4939-3474-4_141

Auer, M. K., Nordenström, A., Lajic, S., & Reisch, N. (2023). Congenital adrenal hyperplasia. *Lancet*, *401*(10372), 227-244. https://doi.org/10.1016/s0140-6736(22)01330-7

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*(2), 144-148. https://doi.org/10.1111/j.1467-9280.2009.02279.x

Auyeung, B., Lombardo, M. V., & Baron-Cohen, S. (2013). Prenatal and postnatal hormone effects on the human brain and cognition. *Pflugers Arch*, *465*(5), 557-571. https://doi.org/10.1007/s00424-013-1268-2

Bailey, A. A., & Hurd, P. L. (2005). Finger length ratio (2D:4D) correlates with physical aggression in men but not in women. *Biological Psychology*, *68*(3), 215-222. https://doi.org/https://doi.org/10.1016/j.biopsycho.2004.05.001

Bakel, H., & Hall, R. A. S. (2020). The Father-infant relationship beyond caregiving sensitivity. *Attach Hum Dev*, *22*(1), 27-31. https://doi.org/10.1080/14616734.2019.1589058

Bakermans-Kranenburg, M. J., Lotz, A., Alyousefi-van Dijk, K., & van IJzendoorn, M. (2019). Birth of a Father: Fathering in the First 1,000 Days. *Child Development Perspectives*, *13*(4), 247-253. https://doi.org/https://doi.org/10.1111/cdep.12347

Bakker, J. (2022). The role of steroid hormones in the sexual differentiation of the human brain. *Journal of Neuroendocrinology*, *34*(2), e13050. https://doi.org/https://doi.org/10.1111/jne.13050

Bandura, A. (1971). *Social Learning Theory*. General Learning Press.

Barbosa, M., Beeghly, M., Gonçalves, J. L., Moreira, J., Tronick, E., & Fuertes, M. (2019). Predicting Patterns of Regulatory Behavior in the Still-Face Paradigm at 3 Months. *Infancy*, *24*(4), 501-525. https://doi.org/https://doi.org/10.1111/infa.12293

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

Barrett, E., Thurston, S. W., Harrington, D., Bush, N. R., Sathyanarayana, S., Nguyen, R., Zavez, A., Wang, C., & Swan, S. (2020). Digit ratio, a proposed marker of the prenatal hormone environment, is not associated with prenatal sex steroids, anogenital distance, or gender-typed play behavior in preschool age children. *J Dev Orig Health Dis*, 1-10. https://doi.org/10.1017/S2040174420001270

Barrett, J., & Fleming, A. S. (2011). Annual Research Review: All mothers are not created equal: neural and psychobiological perspectives on mothering and the importance of individual differences. *J Child Psychol Psychiatry*, *52*(4), 368-397. https://doi.org/10.1111/j.1469-7610.2010.02306.x

Behrendt, H. F., Scharke, W., Herpertz-Dahlmann, B., Konrad, K., & Firk, C. (2019). Like mother, like child? Maternal determinants of children's early social-emotional development. *Infant Mental Health Journal*, *40*(2), 234-247. https://doi.org/https://doi.org/10.1002/imhj.21765

Beijers, R., Breugelmans, S., Brett, B., Willemsen, Y., Bos, P., & de Weerth, C. (2022). Cortisol and testosterone concentrations during the prenatal and postpartum period forecast later caregiving quality in mothers and fathers. *Horm Behav*, *142*, 105177. https://doi.org/10.1016/j.yhbeh.2022.105177

Belsky, J. (1980). MOTHER-INFANT INTERACTION AT HOME AND IN THE LABORATORY - A COMPARATIVE-STUDY. *Journal of Genetic Psychology*, *137*(1), 37-47. https://doi.org/10.1080/00221325.1980.10532800

Beltz, A. M., Demidenko, M. I., Wilson, S. J., & Berenbaum, S. A. (2023). Prenatal androgen influences on the brain: A review, critique, and illustration of research on congenital adrenal hyperplasia. *Journal of Neuroscience Research*, *101*(5), 563-574. https://doi.org/https://doi.org/10.1002/jnr.24900

Beltz, A. M., Kelly, D. P., & Berenbaum, S. A. (2020). Chapter 27 - Sex differences in brain and behavioral development. In J. Rubenstein, P. Rakic, B. Chen, & K. Y. Kwan (Eds.), *Neural Circuit and Cognitive Development (Second Edition)* (pp. 585-638). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-814411-4.00027-5

Berenbaum, S. A., Bryk, K. K., Nowak, N., Quigley, C. A., & Moffat, S. (2009). Fingers as a marker of prenatal androgen exposure. *Endocrinology*, *150*(11), 5119-5124. https://doi.org/10.1210/en.2009-0774

Bilgin, A., & Wolke, D. (2015). Maternal Sensitivity in Parenting Preterm Children: A Meta-analysis. *Pediatrics*, *136*(1), e177-193. https://doi.org/10.1542/peds.2014-3570

Birbaumer, N., & Schmidt, R. F. (2010a). Endokrine Systeme (Hormone). In N. Birbaumer & R. F. Schmidt (Eds.), *Biologische Psychologie* (7th ed., pp. 117-140). Springer. https://doi.org/https://doi.org/10.1007/978-3-540-95938-0

Birbaumer, N., & Schmidt, R. F. (2010b). Psychoneuroendokrinologie. In N. Birbaumer & R. F. Schmidt (Eds.), *Biologische Psychologie* (7th ed., pp. 141-156). Springer. https://doi.org/https://doi.org/10.1007/978-3-540-95938-0

Blair, C., Granger, D., Willoughby, M., & Kivlighan, K. (2006). Maternal sensitivity is related to hypothalamic-pituitary-adrenal axis stress reactivity and regulation in response to emotion challenge in 6-month-old infants. *Ann N Y Acad Sci*, *1094*, 263-267. https://doi.org/10.1196/annals.1376.031

Blair, C., & Ku, S. (2022). A Hierarchical Integrated Model of Self-Regulation [Conceptual Analysis]. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.725828

Bornstein, M. H., & Putnick, D. L. (2007). Chronological age, cognitions, and practices in European American mothers: a multivariate study of parenting. *Dev Psychol*, *43*(4), 850-864. https://doi.org/10.1037/0012-1649.43.4.850

Bornstein, M. H., Putnick, D. L., & Suwalsky, J. T. D. (2019). Mother–infant interactions with firstborns and secondborns: A within-family study of European Americans. *Infant Behavior & Development*, *55*, 100-111. https://doi.org/10.1016/j.infbeh.2019.03.009

Bornstein, M. H., Putnick, D. L., Suwalsky, J. T. D., & Gini, M. (2006). Maternal Chronological Age, Prenatal and Perinatal History, Social Support, and Parenting of Infants. *Child Development*, *77*(4), 875-892. https://doi.org/https://doi.org/10.1111/j.1467-8624.2006.00908.x

Bos, P. A., Hechler, C., Beijers, R., Shinohara, K., Esposito, G., & de Weerth, C. (2018). Prenatal and postnatal cortisol and testosterone are related to parental caregiving quality in fathers, but not in mothers. *Psychoneuroendocrinology*, *97*, 94-103. https://doi.org/10.1016/j.psyneuen.2018.07.013

Bos, P. A., Lesemann, F. H. P., Spencer, H., Stein, D. J., van Honk, J., & Montoya, E. R. (2021). Preliminary data on increased reactivity towards children in distress after testosterone administration in women: A matter of protection? *Biological Psychology*, *165*, 108176. https://doi.org/https://doi.org/10.1016/j.biopsycho.2021.108176

Bowlby, J., & Ainsworth, M. (2000). The Origins of Attachment Theory. In S. Goldberg, R. Muir, & J. Kerr (Eds.), *Attachment Theory. Social Developmental, and Clinical Perspectives* (Vol. 1, pp. 45-84). The Analytic Press, Inc.

Brandt, J. M., Allen, G. A., Haynes, J. L., & Butler, M. G. (1990). NORMATIVE STANDARDS AND COMPARISON OF ANTHROPOMETRIC DATA OF WHITE AND BLACK NEWBORN INFANTS. *Dysmorphol Clin Gen*, *4*(4), 121-137. https://pubmed.ncbi.nlm.nih.gov/27695162

Breedlove, S. M. (2010). Minireview: Organizational hypothesis: instances of the fingerpost. *Endocrinology*, *151*(9), 4116-4122. https://doi.org/10.1210/en.2010-0041

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*(4), 380-386. https://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 radio. *Hum Reprod*, *18*(5), 976-979. https://doi.org/10.1093/humrep/deg198

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*(4), 676-713. https://doi.org/10.1037/0033-295X.106.4.676

Butovskaya, M., Burkova, V., Apalkova, Y., Dronova, D., Rostovtseva, V., Karelin, D., Mkrtchyan, R., Negasheva, M., & Batsevich, V. (2021). Sex, population origin, age and average digit length as predictors of digit ratio in three large world populations. *Sci Rep*, *11*(1), 8157. https://doi.org/10.1038/s41598-021-87394-6

Byrd-Craven, J., Auer, B. J., Granger, D. A., & Massey, A. R. (2012). The father-daughter dance: the relationship between father-daughter relationship quality and daughters' stress response. *J Fam Psychol*, *26*(1), 87-94. https://doi.org/10.1037/a0026588

Califf, R. M. (2018). Biomarker definitions and their applications. *Exp Biol Med (Maywood)*, *243*(3), 213-221. https://doi.org/10.1177/1535370217750088

Calkins, S. D., Propper, C., & Mills-Koonce, W. R. (2013). A biopsychosocial perspective on parenting and developmental psychopathology. *Development and psychopathology*, *25*(4pt2), 1399-1414. https://doi.org/10.1017/S0954579413000680

Carlsen, S. M., Jacobsen, G., & Bjerve, K. S. (2003). Androgen levels in pregnant women decrease with increasing maternal age. *Scandinavian Journal of Clinical and Laboratory Investigation*, *63*(1), 23-26. https://doi.org/10.1080/00365510310000457

Carpenter, T., Grecian, S. M., & Reynolds, R. M. (2017). Sex differences in early-life programming of the hypothalamic–pituitary–adrenal axis in humans suggest increased vulnerability in females: a systematic review. *Journal of Developmental Origins of Health and Disease*, *8*(2), 244-255. https://doi.org/10.1017/S204017441600074X

Chaby, L. E. (2016). Why are there lasting effects from exposure to stress during development? An analysis of current models of early stress. *Physiology & Behavior*, *164*, 164-181. https://doi.org/https://doi.org/10.1016/j.physbeh.2016.05.032

Cherak, S. J., Giesbrecht, G. F., Metcalfe, A., Ronksley, P. E., & Malebranche, M. E. (2018). The effect of gestational period on the association between maternal prenatal salivary cortisol and birth weight: A systematic review and meta-analysis. *Psychoneuroendocrinology*, *94*, 49-62. https://doi.org/https://doi.org/10.1016/j.psyneuen.2018.04.023

Cho, J., Su, X., Phillips, V., & Holditch-Davis, D. (2016). Associations of Maternal and Infant Testosterone and Cortisol Levels With Maternal Depressive Symptoms and Infant Socioemotional Problems. *Biol Res Nurs*, *18*(1), 31-42. https://doi.org/10.1177/1099800415585157

Chrousos, G. P. (2000). Stress, chronic inflammation, and emotional and physical well-being: Concurrent effects and chronic sequelae. *Journal of Allergy and Clinical Immunology*, *106*(5, Supplement), S275-S291. https://doi.org/https://doi.org/10.1067/mai.2000.110163

Cohen-Bendahan, 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*(2), 353-384. https://doi.org/10.1016/j.neubiorev.2004.11.004

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Lawrence Erlbaum Associates.

Cole, L., & Kramer, P. R. (2016). Chapter 3.1 - The Endocrine System. In L. Cole & P. R. Kramer (Eds.), *Human Physiology, Biochemistry and Basic Medicine* (pp. 69-77). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-803699-0.00003-7

Constantinescu, M., & Hines, M. (2012). Relating Prenatal Testosterone Exposure to Postnatal Behavior in Typically Developing Children: Methods and Findings. *Child Dev Perspect*, n/a-n/a. https://doi.org/10.1111/j.1750-8606.2012.00257.x

Coyne, S. M., Manning, J. T., Ringer, L., & Bailey, L. (2007). Directional asymmetry (right–left differences) in digit ratio (2D:4D) predict indirect aggression in women. *Pers Individ Dif*, *43*(4), 865-872. https://doi.org/10.1016/j.paid.2007.02.010

Crnic, K. A., & Greenberg, M. T. (1990). Minor Parenting Stresses with Young Children. *Child Development*, *61*(5), 1628-1637. https://doi.org/https://doi.org/10.1111/j.1467-8624.1990.tb02889.x

D'Anna-Hernandez, K. L., Ross, R. G., Natvig, C. L., & Laudenslager, M. L. (2011). Hair cortisol levels as a retrospective marker of hypothalamic–pituitary axis activity throughout pregnancy: Comparison to salivary cortisol. *Physiology & Behavior*, *104*(2), 348-353. https://doi.org/https://doi.org/10.1016/j.physbeh.2011.02.041

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 [Review]. *Archives of Sexual Behavior*, *49*(2), 373-394. https://doi.org/10.1007/s10508-019-01624-7

de Weerth, C., van Hees, Y., & Buitelaar, J. K. (2003). Prenatal maternal cortisol levels and infant behavior during the first 5 months. *Early Human Development*, *74*(2), 139-151. https://doi.org/https://doi.org/10.1016/S0378-3782(03)00088-4

DeBruine, L. M. (2004). *Autometric Software Measurement of 2D: 4D Ratios*. In https://facelab.org/debruine/Programs/autometric.php

Delemarre-van de Waal, H. A., van Coeverden, S. C., & Rotteveel, J. (2001). Hormonal determinants of pubertal growth. *J Pediatr Endocrinol Metab*, *14 Suppl 6*, 1521-1526.

deMontigny, F., Gervais, C., Pierce, T., & Lavigne, G. (2020). Perceived Paternal Involvement, Relationship Satisfaction, Mothers' Mental Health and Parenting Stress: A Multi-Sample Path Analysis [Original Research]. *Frontiers in Psychiatry*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyt.2020.578682

DiCorcia, J. A., & Tronick, E. (2011). Quotidian resilience: Exploring mechanisms that drive resilience from a perspective of everyday stress and coping. *Neuroscience & Biobehavioral Reviews*, *35*(7), 1593-1602. https://doi.org/https://doi.org/10.1016/j.neubiorev.2011.04.008

Dressler, S. G., & Voracek, M. (2011). No Association Between Two Candidate Markers of Prenatal Sex Hormones: Digit Ratios (2D:4D and Other) and Finger-Ridge Counts [Article]. *Dev Psychobiol*, *53*(1), 69-78. https://doi.org/10.1002/dev.20488

Duden online. (2023). Retrieved September 22, 2023 from https://www.duden.de/rechtschreibung/Hormon

Duthie, L., & Reynolds, R. M. (2013). Changes in the maternal hypothalamic-pituitary-adrenal axis in pregnancy and postpartum: influences on maternal and fetal outcomes. *Neuroendocrinology*, *98*(2), 106-115. https://doi.org/10.1159/000354702

Eichler, A., Kaufmann, F., Titzmann, A., Fasching, P. A., Beckmann, M. W., Gerlach, J., Kratz, O., Moll, G. H., Buchholz, V. N., Kornhuber, J., & Lenz, B. (2023). 2D:4D biomarker reliability and validity in a within-subject pregnancy-childhood-adolescence cohort. *Early Human Development*, *181*, 105776. https://doi.org/https://doi.org/10.1016/j.earlhumdev.2023.105776

Emmanuel, E., & St John, W. (2010). Maternal distress: a concept analysis. *Journal of Advanced Nursing*, *66*(9), 2104-2115. https://doi.org/https://doi.org/10.1111/j.1365-2648.2010.05371.x

Endendijk, J. J., Groeneveld, M. G., Bakermans-Kranenburg, M. J., & Mesman, J. (2016). Gender-Differentiated Parenting Revisited: Meta-Analysis Reveals Very Few Differences in Parental Control of Boys and Girls. *PLoS One*, *11*(7), e0159193. https://doi.org/10.1371/journal.pone.0159193

Enlow, M. B., King, L., Schreier, H. M. C., Howard, J. M., Rosenfield, D., Ritz, T., & Wright, R. J. (2014). Maternal sensitivity and infant autonomic and endocrine stress responses. *Early Human Development*, *90*(7), 377-385. https://doi.org/10.1016/j.earlhumdev.2014.04.007

Ernsten, L., Körner, L. M., Heil, M., Richards, G., & Schaal, N. K. (2021). Investigating the reliability and sex differences of digit lengths, ratios, and hand measures in infants. *Sci Rep*, *11*(1), 10998. https://doi.org/10.1038/s41598-021-89590-w

Ernsten, L., Körner, L. M., Heil, M., & Schaal, N. K. (2024). The association between 2D:4D digit ratio and sex-typed play in children with and without siblings. *Scientific Reports*, *14*(1), 15231. https://doi.org/10.1038/s41598-024-65739-1

Ernsten, L., Körner, L. M., Schaper, M. L., Lawrenz, J., Richards, G., Heil, M., & Schaal, N. K. (2023). The association of prenatal amniotic sex hormones and digit ratio (2D:4D) in children aged 5 to 70

months: A longitudinal study. *PLoS One*, *18*(3), e0282253. https://doi.org/10.1371/journal.pone.0282253

Ernsten, L., Seehagen, S., Nonnenmacher, N., Wolf, O. T., Heil, M., & Schaal, N. K. (accepted for publication after minor revision). *Preliminary findings on the association between maternal salivary and hair cortisol and the mother-infant-interaction during the early postpartum period*. Manuscript accepted for publication in Psychoneuroendocrinology.

Esplen, E. (2009). *Gender and Care: Overview Report*. IDS, BRIDGE Cutting Edge Pack.

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*(1), 101-112. https://doi.org/10.1080/00221325.1989.9914580

Finegood, E. D., Blair, C., Granger, D. A., Hibel, L. C., & Mills-Koonce, R. (2016). Psychobiological influences on maternal sensitivity in the context of adversity. *Dev Psychol*, *52*(7), 1073-1087. https://doi.org/10.1037/dev0000123

Fish, M., & Stifter, C. A. (1993). Mother parity as a main and moderating influence on early mother-infant interaction. *Journal of Applied Developmental Psychology*, *14*(4), 557-572. https://doi.org/https://doi.org/10.1016/0193-3973(93)90007-I

Fleming, A. S., Ruble, D., Krieger, H., & Wong, P. Y. (1997). Hormonal and Experiential Correlates of Maternal Responsiveness during Pregnancy and the Puerperium in Human Mothers. *Hormones and Behavior*, *31*(2), 145-158. https://doi.org/https://doi.org/10.1006/hbeh.1997.1376

Fonseca, C. A. D., Martelli, D. R. B., Maia, C. M. F., Dias, V. O., Carvalho, A. A., & Júnior, H. M. (2022). Digital biomarker 2D:4D as a predictor of cancer: A systematic review. *Early Hum Dev*, *164*, 105521. https://doi.org/10.1016/j.earlhumdev.2021.105521

Frick, M. A., Forslund, T., Fransson, M., Johansson, M., Bohlin, G., & Brocki, K. C. (2018). The role of sustained attention, maternal sensitivity, and infant temperament in the development of early self-regulation. *British Journal of Psychology*, *109*(2), 277-298. https://doi.org/https://doi.org/10.1111/bjop.12266

Galbally, M., Watson, S. J., Lappas, M., de Kloet, E. R., van Rossum, E., Wyrwoll, C., Mark, P., & Lewis, A. J. (2021). Fetal programming pathway from maternal mental health to infant cortisol functioning: The role of placental 11β-HSD2 mRNA expression. *Psychoneuroendocrinology*, *127*, 105197. https://doi.org/https://doi.org/10.1016/j.psyneuen.2021.105197

Galis, F., Ten Broek, C. M., Van Dongen, S., & Wijnaendts, L. C. (2010). Sexual dimorphism in the prenatal digit ratio (2D:4D). *Arch Sex Behav*, *39*(1), 57-62. https://doi.org/10.1007/s10508-009-9485-7

Gettler, L. T., Kuo, P. X., & Bechayda, S. A. (2018). Fatherhood and psychobiology in the Philippines: Perspectives on joint profiles and longitudinal changes of fathers' estradiol and testosterone. *Am J Hum Biol*, *30*(6), e23150. https://doi.org/10.1002/ajhb.23150

Gettler, L. T., McDade, T. W., Feranil, A. B., & Kuzawa, C. W. (2011). Longitudinal evidence that fatherhood decreases testosterone in human males. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(39), 16194-16199. https://doi.org/10.1073/pnas.1105403108

Gianino, A., & Tronick, E. Z. (1988). The mutual regulation model: The infant's self and interactive regulation and coping and defensive capacities. In *Stress and coping across development.* (pp. 47-68). Lawrence Erlbaum Associates, Inc.

Gillam, L., McDonald, R., Ebling, F. J. P., & Mayhew, T. M. (2008). Human 2D (index) and 4D (ring) finger lengths and ratios: cross-sectional data on linear growth patterns, sexual dimorphism and lateral asymmetry from 4 to 60 years of age. *J Anat*, *213*(3), 325-335. https://doi.org/10.1111/j.1469-7580.2008.00940.x

Gilles, M., Otto, H., Wolf, I. A. C., Scharnholz, B., Peus, V., Schredl, M., Sütterlin, M. W., Witt, S. H., Rietschel, M., Laucht, M., & Deuschle, M. (2018). Maternal hypothalamus-pituitary-adrenal (HPA) system activity and stress during pregnancy: Effects on gestational age and infant's anthropometric measures at birth. *Psychoneuroendocrinology*, *94*, 152-161. https://doi.org/https://doi.org/10.1016/j.psyneuen.2018.04.022

Golombok, S., & Fivush, R. (1994). *Gender development*. Cambridge University Press.

Golombok, S., & Hines, M. (2002). Sex differences in social behavior. In *Blackwell handbook of childhood social development.* (pp. 117-136). Blackwell Publishing.

Golombok, S., & Rust, J. (1993). The Pre-School Activities Inventory: A Standardized Assessment of Gender Role in Children. *Psycgoloical Assessment*, *5*(2), 131-136. https://doi.org/10.1037/1040-3590.5.2.131

Goodman, F. R. (2002). Limb malformations and the human HOX genes. *American Journal of Medical Genetics*, *112*(3), 256-265. https://doi.org/https://doi.org/10.1002/ajmg.10776

Grebe, N. M., Sarafin, R. E., Strenth, C. R., & Zilioli, S. (2019). Pair-bonding, fatherhood, and the role of testosterone: A meta-analytic review. *Neuroscience and Biobehavioral Reviews*, *98*, 221-233. https://doi.org/10.1016/j.neubiorev.2019.01.010

Gu, J., & Kanai, R. (2014). What contributes to individual differences in brain structure? [Mini Review]. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00262

Guidi, J., Lucente, M., Sonino, N., & Fava, G. A. (2020). Allostatic Load and Its Impact on Health: A Systematic Review. *Psychotherapy and Psychosomatics*. https://doi.org/10.1159/000510696

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, *34*(2), 163-171. https://doi.org/https://doi.org/10.1016/j.psyneuen.2008.10.026

Hickey, M., Doherty, D. A., Hart, R., Norman, R. J., Mattes, E., Atkinson, H. C., & Sloboda, D. M. (2010). Maternal and umbilical cord androgen concentrations do not predict digit ratio (2D:4D) in girls: A prospective cohort study. *Psychoneuroendocrinology*, *35*(8), 1235-1244. https://doi.org/https://doi.org/10.1016/j.psyneuen.2010.02.013

Hines, M. (2002). 70 - Sexual Differentiation of Human Brain and Behavior. In D. W. Pfaff, A. P. Arnold, S. E. Fahrbach, A. M. Etgen, & R. T. Rubin (Eds.), *Hormones, Brain and Behavior* (pp. 425-462). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012532104-4/50072-X

Hines, M. (2003). Sex steroids and Human Behavior: Prenatal Androgen Exposure and Sex-typical Play Behavior in Children. *Ann. N. Y. Acad. Sci.*, *1007*, 272-282. https://doi.org/10.1196/annals.1286.026

Hines, M. (2015). Gendered development. In R. M. Lerner & M. E. Lamb (Eds.), *Handbook of child development and developmental science* (7th ed., pp. 842-887). Wiley.

Hines, M. (2020a). Human gender development. *Neurosci Biobehav Rev*, *118*, 89-96. https://doi.org/10.1016/j.neubiorev.2020.07.018

Hines, M. (2020b). Neuroscience and Sex/Gender: Looking Back and Forward [Article]. *Journal of Neuroscience*, *40*(1), 37-43. https://doi.org/10.1523/jneurosci.0750-19.2019

Hines, M., Golombok, S., Rust, J., Johnston, K. J., Golding, J., & Team, A. L. S. o. P. a. C. S. (2002). Testosterone during Pregnancy and Gender Role Behavior of Preschool Children: A longitudinal population study. *Child Development*, *73*(6), 1678-1687.

Hines, M., Spencer, D., Kung, K. T., 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. *Curr Opin Neurobiol*, *38*, 69-73. https://doi.org/10.1016/j.conb.2016.02.008

Hönekopp, J. (2012). Digit ratio 2D:4D in relation to autism spectrum disorders, empathizing, and systemizing: a quantitative review. *Autism Res*, *5*(4), 221-230. https://doi.org/10.1002/aur.1230

Hönekopp, J., & Thierfelder, C. (2009). Relationships between digit ratio (2D:4D) and sex-typed play behavior in pre-school children. *Pers Individ Dif*, *47*(7), 706-710. https://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*(5), 619-630. https://doi.org/10.1002/ajhb.21054

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/https://doi.org/10.1016/j.paid.2010.05.003

Horan, H., Cheyney, M., Torres, E. G., Eick, G., Bovbjerg, M., & Snodgrass, J. J. (2022). Maternal hair cortisol concentrations across pregnancy and the early postpartum period in a Puerto Rican sample. *American Journal of Human Biology*, *34*(11), e23718. https://doi.org/https://doi.org/10.1002/ajhb.23718

Hruschak, J. L., Palopoli, A. C., Thomason, M. E., & Trentacosta, C. J. (2022). Maternal-fetal attachment, parenting stress during infancy, and child outcomes at age 3 years. *Infant Ment Health J*, *43*(5), 681-694. https://doi.org/10.1002/imhj.22004

Hsu, H.-C., & Fogel, A. (2003). Stability and Transitions in Mother-Infant Face-to-Face Communication During the First 6 Months: A Microhistorical Approach. *Developmental Psychology*, *39*(6), 1061-1082. https://doi.org/10.1037/0012-1649.39.6.1061

Huber, S. E., Lenz, B., Kornhuber, J., & Müller, C. P. (2017). Prenatal androgen-receptor activity has organizational morphological effects in mice. *PLoS One*, *12*(11), e0188752. https://doi.org/10.1371/journal.pone.0188752

Hughes, C., McHarg, G., & White, N. (2018). Sibling influences on prosocial behavior. *Current Opinion in Psychology*, *20*, 96-101. https://doi.org/https://doi.org/10.1016/j.copsyc.2017.08.015

Hughes, I. A., Davies, J. D., Bunch, T. I., Pasterski, V., Mastroyannopoulou, K., & MacDougall, J. (2012). Androgen insensitivity syndrome. *Lancet*, *380*(9851), 1419-1428. https://doi.org/10.1016/s0140-6736(12)60071-3

Hwang, H., Jo, H. W., Yun, B., Kim, T., Yu, D., Park, J., Kim, D. K., & Woo, S. H. (2014). The Second-to-Fourth Digit Ratio in Cryptorchidism: A Case-Control Study. *Korean J Urol*, *55*(2), 140-144. https://doi.org/10.4111/kju.2014.55.2.140

Jänig, W., & Baron, R. (2019). Hypothalamus. In R. Brandes, F. Lang, & R. F. Schmidt (Eds.), *Physiologie des Menschen: mit Pathophysiologie* (pp. 909-915). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56468-4_72

Journal Citation Reports. (2023a). *PLoS One*. Clarivate. Retrieved November 8, 2023 from https://jcr.clarivate.com/jcr-jp/journal-profile?journal=PLOS%20ONE&year=2022&fromPage=%2Fjcr%2Fhome

Journal Citation Reports. (2023b). *Psychoneuroendocrinology*. Clarivate. Retrieved November 8, 2023 from https://jcr.clarivate.com/jcr-jp/journal-profile?journal=PSYCHONEUROENDOCRINO&year=2022&fromPage=%2Fjcr%2Fhome

Journal Citation Reports. (2023c). *Scientific Reports*. Clarivate. Retrieved November 8, 2023 from https://jcr.clarivate.com/jcr-jp/journal-profile?journal=SCI%20REP-UK&year=2022&fromPage=%2Fjcr%2Fhome

Judd, H. L., Robinson, J. D., Young, P. E., & Jones, O. W. (1976). Amniotic fluid testosterone levels in midpregnancy. *Obstet Gynecol*, *48*(6), 690-692.

Juster, R.-P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews*, *35*(1), 2-16. https://doi.org/https://doi.org/10.1016/j.neubiorev.2009.10.002

Kallak, T. K., Hellgren, C., Skalkidou, A., Sandelin-Francke, L., Ubhayasekhera, K., Bergquist, J., Axelsson, O., Comasco, E., Campbell, R. E., & Sundström Poromaa, I. (2017). Maternal and female fetal testosterone levels are associated with maternal age and gestational weight gain. *European Journal of Endocrinology*, *177*(4), 379-388. https://doi.org/10.1530/eje-17-0207

Kanchan, T., & Krishan, K. (2011). Anthropometry of hand in sex determination of dismembered remains - A review of literature. *J Forensic Leg Med*, *18*(1), 14-17. https://doi.org/10.1016/j.jflm.2010.11.013

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

Khoury, J. E., Bosquet Enlow, M., Patwa, M. C., & Lyons-Ruth, K. (2020). Hair cortisol in pregnancy interacts with maternal depressive symptoms to predict maternal disrupted interaction with her infant at 4 months. *Dev Psychobiol*, *62*(6), 768-782. https://doi.org/10.1002/dev.21950

Kiesow, H., Dunbar, R. I. M., Kable, J. W., Kalenscher, T., Vogeley, K., Schilbach, L., Marquand, A. F., Wiecki, T. V., & Bzdok, D. (2020). 10,000 social brains: Sex differentiation in human brain anatomy. *Science Advances*, *6*(12), eaaz1170. https://doi.org/doi:10.1126/sciadv.aaz1170

Kirschbaum, C., & Hellhammer, D. H. (2008). Salivary Cortisol in Psychobiological Research: An Overview. *Neuropsychobiology*, *22*(3), 150-169. https://doi.org/10.1159/000118611

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (2008). The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, *28*(1-2), 76-81. https://doi.org/10.1159/000119004

Kirschbaum, C., Tietze, A., Skoluda, N., & Dettenborn, L. (2009). Hair as a retrospective calendar of cortisol production-Increased cortisol incorporation into hair in the third trimester of pregnancy. *Psychoneuroendocrinology*, *34*(1), 32-37. https://doi.org/10.1016/j.psyneuen.2008.08.024

Kleine, B., & Rossmanith, W. (2014). Biochemie von Hormonen - Grundlagen. In B. Kleine & W. Rossmanith (Eds.), *Hormone und Hormonsystem - Lehrbuch der Endokrinologie* (Vol. 3., pp. 10-18). Springer-Verlag.

Knickmeyer, R. C., Wheelwright, S., Taylor, K., Raggatt, P., Hackett, G., & Baron-Cohen, S. (2005). Gender-typed play and amniotic testosterone. *Dev Psychol*, *41*(3), 517-528. https://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*(3), 256-263. https://doi.org/10.1016/j.yhbeh.2011.05.009

Knight, E. L., Sarkar, A., Prasad, S., & Mehta, P. H. (2020). Beyond the challenge hypothesis: The emergence of the dual-hormone hypothesis and recommendations for future research. *Hormones and Behavior*, *123*, 104657. https://doi.org/https://doi.org/10.1016/j.yhbeh.2019.104657

Korja, R., Latva, R., & Lehtonen, L. (2012). The effects of preterm birth on mother–infant interaction and attachment during the infant's first two years [https://doi.org/10.1111/j.1600-0412.2011.01304.x]. *Acta Obstetricia et Gynecologica Scandinavica*, *91*(2), 164-173. https://doi.org/https://doi.org/10.1111/j.1600-0412.2011.01304.x

Körner, L. M., Schaper, M. L., Pause, B. M., & Heil, M. (2020). Parent-Reports of Sex-Typed Play Preference in Preschool Children: Relationships to 2D:4D Digit Ratio and Older Siblings' Sex. *Arch Sex Behav*, *49*(7), 2715-2724. https://doi.org/10.1007/s10508-020-01662-6

Kota, S. K., Gayatri, K., Jammula, S., Meher, L. K., Kota, S. K., Krishna, S. V. S., & Modi, K. D. (2013). Fetal endocrinology. *Indian Journal of Endocrinology and Metabolism*, *17*(4), 568-579. https://doi.org/10.4103/2230-8210.113722

Kralik, M., Hupková, A., Zeman, T., Hložek, M., Hlaváček, L., Slováčková, M., Čuta, M., & Urbanová, P. (2019). Family effects on the digit ratio (2D:4D): The role of the interbirth interval. *American Journal of Human Biology*, *31*. https://doi.org/10.1002/ajhb.23260

Králík, M., Ingrová, P., Kozieł, S., Hupková, A., & Klíma, O. (2017). Overall trends vs. individual trajectories in the second-to-fourth digit (2D:4D) and metacarpal (2M:4M) ratios during puberty and adolescence. *American Journal of Physical Anthropology*, *162*(4), 641-656. https://doi.org/https://doi.org/10.1002/ajpa.23153

Krishan, K., Kanchan, T., & Sharma, A. (2011). Sex determination from hand and foot dimensions in a North Indian population. *J Forensic Sci*, *56*(2), 453-459. https://doi.org/10.1111/j.1556-4029.2010.01652.x

Kuchirko, Y., Bennet, A., Halim, M. L., Costanzo, P., & Ruble, D. (2021). The influence of siblings on ethnically diverse children's gender typing across early development. *Developmental Psychology*, *57*(5), 771-782. https://doi.org/10.1037/dev0001173

Kudielka, B. M., & Kirschbaum, C. (2005). Sex differences in HPA axis responses to stress: a review. *Biol Psychol*, *69*(1), 113-132. https://doi.org/10.1016/j.biopsycho.2004.11.009

Kuijper, E. A. M., Ket, J. C. F., Caanen, M. R., & Lambalk, C. B. (2013). Reproductive hormone concentrations in pregnancy and neonates: a systematic review. *Reprod Biomed Online*, *27*(1), 33-63. https://doi.org/https://doi.org/10.1016/j.rbmo.2013.03.009

Kumar, S., Voracek, M., & Singh, M. (2017). Sexual Dimorphism in Digit Ratios Derived from Dorsal Digit Length among Adults and Children. *Front Endocrinol (Lausanne)*, *8*, 41. https://doi.org/10.3389/fendo.2017.00041

Kuo, P. X., Braungart-Rieker, J. M., Burke Lefever, J. E., Sarma, M. S., O'Neill, M., & Gettler, L. T. (2018). Fathers' cortisol and testosterone in the days around infants' births predict later paternal

involvement. *Hormones and Behavior*, *106*, 28-34. https://doi.org/https://doi.org/10.1016/j.yhbeh.2018.08.011

Kuzawa, C. W. (1998). Adipose tissue in human infancy and childhood: An evolutionary perspective. *Am J Phys Anthropol*, *107*(S27), 177-209. https://doi.org/10.1002/(SICI)1096-8644(1998)107:27+<177::AID-AJPA7>3.0.CO;2-B

Kyriakidis, I. (2021). Data regarding 2D:4D and other digit ratios in Greek population. *Data in Brief*, *34*, 106724. https://doi.org/https://doi.org/10.1016/j.dib.2021.106724

Kyriakidis, I., & Papaioannidou, P. (2008). Epidemiologic study of the sexually dimorphic second to fourth digit ratio (2D:4D) and other finger ratios in Greek population. *Coll Antropol*, *32*(4), 1093-1098.

Lamminmaki, A., Hines, M., Kuiri-Hanninen, T., Kilpelainen, L., Dunkel, L., & Sankilampi, U. (2012). Testosterone measured in infancy predicts subsequent sex-typed behavior in boys and in girls. *Horm Behav*, *61*(4), 611-616. https://doi.org/10.1016/j.yhbeh.2012.02.013

Lang, F., & Föller, M. (2019). Nebennierenrindenhormone. In R. Brandes, F. Lang, & R. F. Schmidt (Eds.), *Phsiologie des Menschen. Springer-Lehrbuch* (pp. 943-950). Springer-Verlag.

Laurent, H. K., Harold, G. T., Leve, L., Shelton, K. H., & Van Goozen, S. H. M. (2016). Understanding the unfolding of stress regulation in infants. *Development and Psychopathology*, *28*(4pt2), 1431-1440. https://doi.org/10.1017/S0954579416000171

LeMoult, J., Humphreys, K. L., Tracy, A., Hoffmeister, J.-A., Ip, E., & Gotlib, I. H. (2020). Meta-analysis: Exposure to Early Life Stress and Risk for Depression in Childhood and Adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *59*(7), 842-855. https://doi.org/https://doi.org/10.1016/j.jaac.2019.10.011

Lewis, A. J., Austin, E., Knapp, R., Vaiano, T., & Galbally, M. (2015). Perinatal Maternal Mental Health, Fetal Programming and Child Development. *Healthcare (Basel)*, *3*(4), 1212-1227. https://doi.org/10.3390/healthcare3041212

Loehlin, J. C., & Medland, S. E. (2008). Multivariate Genetic Analyses of the 2D:4D Ratio: Examining the Effects of Hand and Measurement Technique in Data from 757 Twin Families. *Twin Research and Human Genetics*, *11*(3), 335-341. https://doi.org/10.1375/twin.11.3.335

Loehlin, J. C., Medland, S. E., & Martin, N. G. (2009). Relative finger lengths, sex differences, and psychological traits. *Arch Sex Behav*, *38*(2), 298-305. https://doi.org/10.1007/s10508-007-9303-z

Lolli, L., Batterham, A. M., Kratochvíl, L., Flegr, J., Weston, K. L., & Atkinson, G. (2017). A comprehensive allometric analysis of 2nd digit length to 4th digit length in humans. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1857), 20170356. https://doi.org/doi:10.1098/rspb.2017.0356

Lonsdorf, E. V. (2017). Sex differences in nonhuman primate behavioral development. *Journal of Neuroscience Research*, *95*(1-2), 213-221. https://doi.org/10.1002/jnr.23862

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, *10*(6), 434-445. https://doi.org/10.1038/nrn2639

Lupu, D. C., Monedero, I., Rodriguez-Ruiz, C., Pita, M., & Turiegano, E. (2023). In support of 2D:4D: More data exploring its conflicting results on handedness, sexual orientation and sex differences. *PLoS One*, *18*(8), e0280514. https://doi.org/10.1371/journal.pone.0280514

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*(1-2), 23-28. https://doi.org/10.1016/j.earlhumdev.2003.12.002

Malas, M. A., Dogan, S., Evcil, E. H., & Desdicioglu, K. (2006). Fetal development of the hand, digits and digit ratio (2D:4D). *Early Hum Dev*, *82*(7), 469-475. https://doi.org/10.1016/j.earlhumdev.2005.12.002

Manning, J., Cook, C., & Crewther, B. (2019). Digit ratio (2D:4D) and testosterone supplementation. *Early Hum Dev*, *139*, 104843. https://doi.org/10.1016/j.earlhumdev.2019.104843

Manning, J., Kilduff, L., Cook, C., Crewther, B., & Fink, B. (2014). Digit Ratio (2D:4D): A Biomarker for Prenatal Sex Steroids and Adult Sex Steroids in Challenge Situations [Hypothesis and Theory]. *Frontiers in Endocrinology*, *5*. https://doi.org/10.3389/fendo.2014.00009

Manning, J., Scutt, D., Wilson, J., & Lewis-Jones, D. (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. https://doi.org/10.1093/humrep/13.11.3000

Manning, J. T. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. Rutgers University Press.

Manning, J. T. (2012). Sex Differences and Age Changes in Digit Ratios: Implications for the Use of Digit Ratios in Medicine and Biology. In V. R. Preedy (Ed.), *Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease* (pp. 841-851). Springer New York. https://doi.org/10.1007/978-1-4419-1788-1_48

Manning, J. T., Churchill, A. J. G., & Peters, M. (2007). The Effects of Sex, Ethnicity, and Sexual Orientation on Self-Measured Digit Ratio (2D:4D). *Archives of Sexual Behavior*, *36*(2), 223-233. https://doi.org/10.1007/s10508-007-9171-6

Manning, J. T., & Fink, B. (2018). Sexual dimorphism in the ontogeny of second (2D) and fourth (4D) digit lengths, and digit ratio (2D:4D). *Am J Hum Biol*, *30*(4), e23138. https://doi.org/10.1002/ajhb.23138

Manning, J. T., & Fink, B. (2023). Digit ratio (2D:4D) and its relationship to foetal and maternal sex steroids: A mini-review. *Early Hum Dev*, *183*, 105799. https://doi.org/10.1016/j.earlhumdev.2023.105799

Marin, M.-F., Lord, C., Andrews, J., Juster, R.-P., Sindi, S., Arsenault-Lapierre, G., Fiocco, A. J., & Lupien, S. J. (2011). Chronic stress, cognitive functioning and mental health. *Neurobiology of Learning and Memory*, *96*(4), 583-595. https://doi.org/https://doi.org/10.1016/j.nlm.2011.02.016

Marley, C. L., Pollard, T. M., Barton, R. A., & Street, S. E. (2022). A systematic review of sex differences in rough and tumble play across non-human mammals. *Behavioral Ecology and Sociobiology*, *76*(12), 158. https://doi.org/10.1007/s00265-022-03260-z

Marshall, W. A. (1975). Growth and sexual maturation in normal puberty. *Clinics in Endocrinology and Metabolism*, *4*(1), 3-25. https://doi.org/https://doi.org/10.1016/S0300-595X(75)80032-6

Matzdorff, I. (1967). Metrische und allometrische Wachstumsuntersuchungen an der menschlichen Hand. *Z Morphol Anthropol*, *59*(2), 158-184. http://www.jstor.org/stable/25755113

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424-453. https://doi.org/10.1037/0033-2909.126.3.424

McFadden, D., & Shubel, E. (2002). Relative lengths of fingers and toes in human males and females. *Horm Behav*, *42*(4), 492-500. https://doi.org/10.1006/hbeh.2002.1833

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

McIntyre, M. H., Cohn, B. A., & Ellison, P. T. (2006). Sex dimorphism in digital formulae of children. *Am J Phys Anthropol*, *129*(1), 143-150. https://doi.org/https://doi.org/10.1002/ajpa.20240

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 Lond B Biol Sci*, *272*(1571), 1473-1479. https://doi.org/doi:10.1098/rspb.2005.3100

Mehta, P. H., & Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior*, *58*(5), 898-906. https://doi.org/https://doi.org/10.1016/j.yhbeh.2010.08.020

Mehta, P. H., & Prasad, S. (2015). The dual-hormone hypothesis: a brief review and future research agenda. *Current Opinion in Behavioral Sciences*, *3*, 163-168. https://doi.org/https://doi.org/10.1016/j.cobeha.2015.04.008

Meyer, J., Novak, M., Hamel, A., & Rosenberg, K. (2014). Extraction and analysis of cortisol from human and monkey hair. *J Vis Exp*(83), e50882. https://doi.org/10.3791/50882

Meyer, J. S., & Novak, M. A. (2012). Minireview: Hair cortisol: a novel biomarker of hypothalamic-pituitary-adrenocortical activity. *Endocrinology*, *153*(9), 4120-4127. https://doi.org/10.1210/en.2012-1226

Mikac, U., Buško, V., Sommer, W., & Hildebrandt, A. (2016). Analysis of different sources of measurement error in determining second-to-fourth digit ratio, a potential indicator of perinatal sex hormones exposure. *Annu Rev Psychol*, *23*, 39-49. https://doi.org/10.21465/rp0023.0004

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin*, *133*(1), 25-45. https://doi.org/10.1037/0033-2909.133.1.25

Mills-Koonce, W. R., Propper, C., Gariepy, J.-L., Barnett, M., Moore, G. A., Calkins, S., & Cox, M. J. (2009). Psychophysiological correlates of parenting behavior in mothers of young children. *Developmental Psychobiology*, *51*(8), 650-661. https://doi.org/https://doi.org/10.1002/dev.20400

Mitsui, T., Araki, A., Miyashita, C., Ito, S., Ikeno, T., Sasaki, S., Kitta, T., Moriya, K., Cho, K., Morioka, K., Kishi, R., Shinohara, N., Takeda, M., & Nonomura, K. (2016). The Relationship between the Second-to-Fourth Digit Ratio and Behavioral Sexual Dimorphism in School-Aged Children. *PLoS One*, *11*(1), e0146849. https://doi.org/10.1371/journal.pone.0146849

Mueller, I., Snidman, N., DiCorcia, J. A., & Tronick, E. (2021). Acute Maternal Stress Disrupts Infant Regulation of the Autonomic Nervous System and Behavior: A CASP Study [Original Research]. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.714664

Mughal, M. K., Giallo, R., Arnold, P., Benzies, K., Kehler, H., Bright, K., & Kingston, D. (2018). Trajectories of maternal stress and anxiety from pregnancy to three years and child development at 3 years of age: Findings from the All Our Families (AOF) pregnancy cohort. *Journal of Affective Disorders*, *234*, 318-326. https://doi.org/https://doi.org/10.1016/j.jad.2018.02.095

Nassar, G. N., & Leslie, S. W. (2022). *Physiology, Testosterone*. StatPearls Publishing, Treasure Island (FL). http://europepmc.org/abstract/MED/30252384

Nave, G., Koppin, C. M., Manfredi, D., Richards, G., Watson, S. J., Geffner, M. E., Yong, J. E., Kim, R., Ross, H. M., Serrano-Gonzalez, M., & Kim, M. S. (2021). No evidence for a difference in 2D:4D ratio between youth with elevated prenatal androgen exposure due to congenital adrenal hyperplasia and controls. *Horm Behav*, *128*, 104908. https://doi.org/10.1016/j.yhbeh.2020.104908

Nelson, A. M. (2003). Transition to Motherhood. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, *32*(4), 465-477. https://doi.org/https://doi.org/10.1177/0884217503255199

Nicolaides, K. H. (2005). First-trimester screening for chromosomal abnormalities. *Semin Perinatol*, *29*(4), 190-194. https://doi.org/10.1053/j.semperi.2005.06.001

Nolvi, S., Karlsson, L., Bridgett, D. J., Korja, R., Huizink, A. C., Kataja, E.-L., & Karlsson, H. (2016). Maternal prenatal stress and infant emotional reactivity six months postpartum. *Journal of Affective Disorders*, *199*, 163-170. https://doi.org/https://doi.org/10.1016/j.jad.2016.04.020

Nystrom-Hansen, M., Andersen, M. S., Khoury, J. E., Davidsen, K., Gumley, A., Lyons-Ruth, K., MacBeth, A., & Harder, S. (2019). Hair cortisol in the perinatal period mediates associations between maternal adversity and disrupted maternal interaction in early infancy [Article]. *Developmental Psychobiology*, *61*(4), 543-556. https://doi.org/10.1002/dev.21833

Öhman, L., Bergdahl, J., Nyberg, L., & Nilsson, L.-G. (2007). Longitudinal analysis of the relation between moderate long-term stress and health. *Stress and Health*, *23*(2), 131-138. https://doi.org/https://doi.org/10.1002/smi.1130

Pasanen, B. E., Tomkinson, J. M., Dufner, T. J., Park, C. W., Fitzgerald, J. S., & Tomkinson, G. R. (2021). The relationship between digit ratio (2D:4D) and muscular fitness: A systematic review and meta-analysis. *Am J Hum Biol*, e23657. https://doi.org/10.1002/ajhb.23657

Penny, R., Parlow, A. F., & Frasier, S. D. (1979). Testosterone and estradiol concentrations in paired maternal and cord sera and their correlation with the concentration of chorionic gonadotropin. *Pediatrics*, *64*(5), 604-608.

Phoenix, C. H., Goy, R. W., Gerall, A. A., & Young, W. C. (1959). Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology*, *65*, 369-382. https://doi.org/10.1210/endo-65-3-369

Pochigaeva, K., Druzhkova, T., Yakovlev, A., Onufriev, M., Grishkina, M., Chepelev, A., Guekht, A., & Gulyaeva, N. (2017). Hair cortisol as a marker of hypothalamic-pituitary-adrenal Axis activity in female patients with major depressive disorder. *Metabolic Brain Disease*, *32*(2), 577-583. https://doi.org/10.1007/s11011-017-9952-0

Provenzi, L., Fumagalli, M., Giorda, R., Morandi, F., Sirgiovanni, I., Pozzoli, U., Mosca, F., Borgatti, R., & Montirosso, R. (2017). Maternal Sensitivity Buffers the Association between SLC6A4

Methylation and Socio-Emotional Stress Response in 3-Month-Old Full Term, but not very Preterm Infants [Original Research]. *Frontiers in Psychiatry*, *8*. https://doi.org/10.3389/fpsyt.2017.00171

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, *28*(7), 916-931. https://doi.org/10.1016/s0306-4530(02)00108-7

Raziye, D., Ceren, U., Kadir, D., Osman, S., & Mehmed Ali, M. (2016). A Radiological Investigation on the Hand Development in Human Fetuses Throughout the Fetal Period and an Evaluation Performed in Terms of its Clinical Importance Hand Development. *International Journal of Morphology*, *34*, 1539-1552. https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0717-95022016000400057&nrm=iso

Reck, C., Noe, D., & Cenciotti, F. (2009). Infant and Caregiver Engagement Phases (ICEP): Heidelberg Version. In U. o. Heidelberg (Ed.). Heidelberg, Germany.

Richards, G., Aydin, E., Tsompanidis, A., Padaigaitė, E., Austin, T., Allison, C., Holt, R., & Baron-Cohen, S. (2022). Digit ratio (2D:4D) and maternal testosterone-to-estradiol ratio measured in early pregnancy. *Sci Rep*, *12*(1), 13586. https://doi.org/10.1038/s41598-022-17247-3

Richards, G., Bellin, W., & Davies, W. (2017). Familial digit ratio (2D:4D) associations in a general population sample from Wales. *Early Hum Dev*, *112*, 14-19. https://doi.org/https://doi.org/10.1016/j.earlhumdev.2017.06.006

Richards, G., Browne, W., & Constantinescu, M. (2020a). Digit ratio (2D:4D) and amniotic testosterone and estradiol: An attempted replication of Lutchmaya et al. (2004). *J Dev Orig Health Dis*. https://doi.org/10.1101/2020.07.10.197269

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

Richards, G., Gomes, M., & Ventura, T. (2019). Testosterone measured from amniotic fluid and maternal plasma shows no significant association with directional asymmetry in newborn digit ratio (2D:4D). *J Dev Orig Health Dis*, *10*(3), 362-367. https://doi.org/10.1017/S2040174418000752

Richter, N., Bondü, R., & Trommsdorff, G. (2022). Linking transition to motherhood to parenting, children's emotion regulation, and life satisfaction: A longitudinal study. *Journal of Family Psychology*, *36*, 291-300. https://doi.org/10.1037/fam0000868

Robinson, J. D., Judd, H. L., Young, P. E., Jones, O. W., & Yen, S. S. (1977). Amniotic fluid androgens and estrogens in midgestation. *J Clin Endocrinol Metab*, *45*(4), 755-761. https://doi.org/10.1210/jcem-45-4-755

Roby, E., Piccolo, L. R., Gutierrez, J., Kesoglides, N. M., Raak, C. D., Mendelsohn, A. L., & Canfield, C. F. (2021). Father involvement in infancy predicts behavior and response to chronic stress in middle childhood in a low-income Latinx sample. *Dev Psychobiol*, *63*(5), 1449-1465. https://doi.org/10.1002/dev.22081

Rogol, A. D., Roemmich, J. N., & Clark, P. A. (2002). Growth at puberty. *Journal of Adolescent Health*, *31*(6, Supplement), 192-200. https://doi.org/https://doi.org/10.1016/S1054-139X(02)00485-8

Rohleder, N. (2019). Stress and inflammation – The need to address the gap in the transition between acute and chronic stress effects. *Psychoneuroendocrinology*, *105*, 164-171. https://doi.org/https://doi.org/10.1016/j.psyneuen.2019.02.021

Rominov, H., Giallo, R., & Whelan, T. A. (2016). Fathers' postnatal distress, parenting self-efficacy, later parenting behavior, and children's emotional-behavioral functioning: A longitudinal study. *J Fam Psychol*, *30*(8), 907-917. https://doi.org/10.1037/fam0000216

Rust, J., Golombok, S., Hines, M., Johnston, K., & Golding, J. (2000). The Role of Brothers and Sisters in the Gender Development of Preschool Children. *Journal of Experimental Child Psychology*, *77*(4), 292-303. https://doi.org/https://doi.org/10.1006/jecp.2000.2596

Saino, N., Leoni, B., & Romano, M. (2006). Human digit ratios depend on birth order and sex of older siblings and predict maternal fecundity. *Behavioral Ecology and Sociobiology*, *60*(1), 34-45. https://doi.org/10.1007/s00265-005-0137-1

Saxbe, D., Rossin-Slater, M., & Goldenberg, D. (2018). The transition to parenthood as a critical window for adult health. *Am Psychol*, *73*(9), 1190-1200. https://doi.org/10.1037/amp0000376

Saxbe, D. E., Edelstein, R. S., Lyden, H. M., Wardecker, B. M., Chopik, W. J., & Moors, A. C. (2017). Fathers' decline in testosterone and synchrony with partner testosterone during pregnancy predicts greater postpartum relationship investment. *Hormones and Behavior*, *90*, 39-47. https://doi.org/https://doi.org/10.1016/j.yhbeh.2016.07.005

Saxbe, D. E., Margolin, G., Spies Shapiro, L., Ramos, M., Rodriguez, A., & Iturralde, E. (2014). Relative influences: Patterns of HPA axis concordance during triadic family interaction. *Health Psychology*, *33*(3), 273-281. https://doi.org/10.1037/a0033509

Schore, A. N. (2001). Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Mental Health Journal*, *22*(1-2), 7-66. https://doi.org/https://onlinelibrary.wiley.com/doi/10.1002/1097-0355(200101/04)22:1%3C7::AID-IMHJ2%3E3.0.CO;2-N

Schwabe, L., & Wolf, O. T. (2013). Stress and multiple memory systems: from 'thinking' to 'doing'. *Trends Cogn Sci*, *17*(2), 60-68. https://doi.org/10.1016/j.tics.2012.12.001

Shin, H., Park, Y.-J., Ryu, H., & Seomun, G.-A. (2008). Maternal sensitivity: a concept analysis. *Journal of Advanced Nursing*, *64*(3), 304-314. https://doi.org/https://doi.org/10.1111/j.1365-2648.2008.04814.x

Short, S. J., Stalder, T., Marceau, K., Entringer, S., Moog, N. K., Shirtcliff, E. A., Wadhwa, P. D., & Buss, C. (2016). Correspondence between hair cortisol concentrations and 30-day integrated daily salivary and weekly urinary cortisol measures. *Psychoneuroendocrinology*, *71*, 12-18. https://doi.org/https://doi.org/10.1016/j.psyneuen.2016.05.007

Singh Solorzano, C., Serwinski, B., Grano, C., & Steptoe, A. (2023). Longitudinal association between saliva and hair cortisol concentration: A systematic comparison. *Psychoneuroendocrinology*, *156*, 106340. https://doi.org/https://doi.org/10.1016/j.psyneuen.2023.106340

Sisk, C., Lonstein, J. S., & Gore, A. C. (2013). Critical Periods During Development: Hormonal Influences on Neurobehavioral Transitions Across the Life Span. In D. W. Pfaff (Ed.), *Neuroscience in the*

21st Century: From Basic to Clinical (pp. 1715-1752). Springer New York. https://doi.org/10.1007/978-1-4614-1997-6_61

Sorokowski, P., & Kowal, M. (2024). Relationship between the 2D:4D and prenatal testosterone, adult level testosterone, and testosterone change: Meta-analysis of 54 studies. *Am J Biol Anthropol*, *183*(1), 20-38. https://doi.org/10.1002/ajpa.24852

Spencer, D., Pasterski, V., 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/https://doi.org/10.1016/j.yhbeh.2020.104889

Stenstrom, E., Saad, G., Nepomuceno, M. V., & Mendenhall, Z. (2011). Testosterone and domain-specific risk: Digit ratios (2D:4D and rel2) as predictors of recreational, financial, and social risk-taking behaviors. *Pers Individ Dif*, *51*(4), 412-416. https://doi.org/10.1016/j.paid.2010.07.003

Suchonova, M., Borbelyova, V., Renczes, E., Konecna, B., Vlkova, B., Hodosy, J., Ostatnikova, D., & Celec, P. (2019). Does the 2nd and 4th digit ratio reflect prenatal androgen exposure? *Bratisl Lek Listy*, *120*(9), 703-710. https://doi.org/10.4149/bll_2019_119

Suhr, F. (2019). *Eltern leiden unter Dauerstress [Digitales Bild]*. Retrieved 24 November from https://de.statista.com/infografik/20250/umfrage-unter-eltern-zum-stress-im-alltag/

Swift-Gallant, A., Di Rita, V., Coome, L. A., & Monks, D. A. (2021). Overexpression of Androgen Receptors Masculinizes 2D:4D Digit Ratios in Mice. *Arch Sex Behav*. https://doi.org/10.1007/s10508-021-02161-y

Swift-Gallant, A., & Monks, D. A. (2017). Androgenic mechanisms of sexual differentiation of the nervous system and behavior. *Front Neuroendocrinol*, *46*, 32-45. https://doi.org/10.1016/j.yfrne.2017.04.003

Takegata, M., Matsunaga, A., Ohashi, Y., Toizumi, M., Yoshida, L. M., & Kitamura, T. (2021). Prenatal and Intrapartum Factors Associated With Infant Temperament: A Systematic Review [Systematic Review]. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.609020

Tal, R., & Taylor, H. S. (2021). *Endocrinology of Pregnancy*. South Dartmouth (MA): MDText.com, Inc.; 2000-. Retrieved 06.03.2023 from https://www.ncbi.nlm.nih.gov/books/NBK278962/

Talarovičová, A., Kršková, L., & Blažeková, J. (2009). Testosterone enhancement during pregnancy influences the 2D:4D ratio and open field motor activity of rat siblings in adulthood. *Horm Behav*, *55*(1), 235-239. https://doi.org/10.1016/j.yhbeh.2008.10.010

Tarullo, A. R., St John, A. M., & Meyer, J. S. (2017). Chronic stress in the mother-infant dyad: Maternal hair cortisol, infant salivary cortisol and interactional synchrony. *Infant behavior & development*, *47*, 92-102. https://doi.org/10.1016/j.infbeh.2017.03.007

Teatero, M. L., & Netley, C. (2013). A critical review of the research on the extreme male brain theory and digit ratio (2D:4D). *J Autism Dev Disord*, *43*(11), 2664-2676. https://doi.org/10.1007/s10803-013-1819-6

Thompson, L. A., & Trevathan, W. R. (2008). Cortisol reactivity, maternal sensitivity, and learning in 3-month-old infants. *Infant Behavior & Development*, *31*(1), 92-106. https://doi.org/10.1016/j.infbeh.2007.07.007

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

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. https://doi.org/10.1016/j.yhbeh.2005.05.023

Troisi, R., Potischman, N., Roberts, J. M., Harger, G., Markovic, N., Cole, B., Lykins, D., Siiteri, P., & Hoover, R. N. (2003). Correlation of Serum Hormone Concentrations in Maternal and Umbilical Cord Samples1. *Cancer Epidemiology, Biomarkers & Prevention*, *12*(5), 452-456.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *J Am Acad Child Psychiatry*, *17*(1), 1-13. https://doi.org/10.1016/s0002-7138(09)62273-1

Tronick, E., Mueller, I., DiCorcia, J., Hunter, R., & Snidman, N. (2021). A Caretaker Acute Stress Paradigm: Effects on behavior and physiology of caretaker and infant. *Developmental Psychobiology*, *63*(2), 237-246. https://doi.org/https://doi.org/10.1002/dev.21974

Tronick, E. Z., & Gianino, A. (1986). Interactive mismatch and repair: Challenges to the coping infant. *Zero to Three*, *6*(3), 1-6.

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/https://doi.org/10.1016/j.yhbeh.2004.06.010

van de Beek, C., van Goozen, S. H., 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*(1), 6-15. https://doi.org/10.1007/s10508-007-9291-z

Van den Bergh, B. R. H., van den Heuvel, M. I., Lahti, M., Braeken, M., de Rooij, S. R., Entringer, S., Hoyer, D., Roseboom, T., Räikkönen, K., King, S., & Schwab, M. (2020). Prenatal developmental origins of behavior and mental health: The influence of maternal stress in pregnancy. *Neuroscience & Biobehavioral Reviews*, *117*, 26-64. https://doi.org/https://doi.org/10.1016/j.neubiorev.2017.07.003

van den Heuvel, M. I. (2022). From the Womb into the World: Protecting the Fetal Brain from Maternal Stress During Pregnancy. *Policy Insights from the Behavioral and Brain Sciences*, *9*(1), 96-103. https://doi.org/10.1177/23727322211068024

van Hemmen, J., Cohen-Kettenis, P. T., Steensma, T. D., Veltman, D. J., & Bakker, J. (2017). Do sex differences in CEOAEs and 2D:4D ratios reflect androgen exposure? A study in women with complete androgen insensitivity syndrome. *Biol Sex Differ*, *8*, 11. https://doi.org/10.1186/s13293-017-0132-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*(2), 107-112. https://doi.org/10.1016/j.earlhumdev.2012.08.009

Voorthuis, A., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2019). Testosterone reactivity to infant crying and caregiving in women: The role of oral contraceptives and basal cortisol. *Infant Behavior and Development*, *56*, 101191. https://doi.org/https://doi.org/10.1016/j.infbeh.2017.08.002

Voracek, M., & Dressler, S. G. (2007). Digit Ratio (2D:4D) in Twins: Heritability Estimates and Evidence for a Masculinized Trait Expression in Women from opposite-Sex Pairs. *Psychological Reports*, *100*(1), 115-126. https://doi.org/10.2466/pr0.100.1.115-126

Voracek, M., & Dressler, S. G. (2009). Brief communication: Familial resemblance in digit ratio (2D:4D). *American Journal of Physical Anthropology*, *140*(2), 376-380. https://doi.org/https://doi.org/10.1002/ajpa.21105

Wallen, K. (2009). Does Finger Fat Produce Sex Differences in Second to Fourth Digit Ratios? *Endocrinology*, *150*(11), 4819-4822. https://doi.org/10.1210/en.2009-0986

Weinberg, M., Tronick, E., Cohn, J., & Olson, K. (1999). Gender differences in emotional expressivity and self-regulation during early infancy. *Developmental Psychology*, *35*, 175-188. https://doi.org/10.1037//0012-1649.35.1.175

Williams, T. J., Pepitone, M. E., Christensen, S. E., Cooke, B. M., Huberman, A. D., Breedlove, N. J., Breedlove, T. J., Jordan, C. L., & Breedlove, S. M. (2000). Finger-length ratios and sexual orientation. *Nature*, *404*(6777), 455-456. https://doi.org/10.1038/35006555

Wong, W. I., & Hines, M. (2016). Interpreting digit ratio (2D:4D)-behavior correlations: 2D:4D sex difference, stability, and behavioral correlates and their replicability in young children [Article]. *Horm Behav*, *78*, 86-94. https://doi.org/10.1016/j.yhbeh.2015.10.022

Young, E., Doom, J., Farrell, A., Carlson, E., Englund, M., Miller, G., Gunnar, M., Roisman, G., & Simpson, J. (2020). Life stress and cortisol reactivity: An exploratory analysis of the effects of stress

exposure across life on HPA-axis functioning. *Development and psychopathology*, 1-12. https://doi.org/10.1017/S0954579419001779

Zegiob, L. E., Arnold, S., & Forehand, R. (1975). An Examination of Observer Effects in Parent-Child Interactions. *Child Development*, *46*(2), 509-512. https://doi.org/10.2307/1128149

Zhang, Q., Chen, Z., Chen, S., Yu, T., Wang, J., Wang, W., & Deng, H. (2018). Correlations of hair level with salivary level in cortisol and cortisone. *Life Sciences*, *193*, 57-63. https://doi.org/https://doi.org/10.1016/j.lfs.2017.11.037

Zheng, Z., & Cohn, M. J. (2011). Developmental basis of sexually dimorphic digit ratios. *Proc Natl Acad Sci U S A*, *108*(39), 16289-16294. https://doi.org/10.1073/pnas.1108312108

Zijlmans, M. A. C., Riksen-Walraven, J. M., & de Weerth, C. (2015). Associations between maternal prenatal cortisol concentrations and child outcomes: A systematic review. *Neuroscience & Biobehavioral Reviews*, *53*, 1-24. https://doi.org/https://doi.org/10.1016/j.neubiorev.2015.02.015

Declaration

Affidavit according to §5 of the doctoral regulations of the Faculty of Mathematics and Natural Sciences of Heinrich Heine University Düsseldorf from 15.06.2018:

I declare under oath that I have produced my thesis independently and without any undue assistance by third parties under consideration of the 'Principles for the Safeguarding of Good Scientific Practice at Heinrich Heine University Düsseldorf'.

Furthermore, I declare that my thesis has not been submitted to any other faculty as a dissertation in the form presented or in a similar form and that I have not made any unsuccessful attempts obtaining a doctorate to date.

Düsseldorf, den _____ _____

(Datum) (Unterschrift)

Danksagung

Die letzten Jahre und auch die Fertigstellung dieser Dissertation wären nicht möglich gewesen ohne die Hilfe und Unterstützung zahlreicher Personen.

An erster Stelle möchte ich Nora Schaal und Martin Heil danken, die mir überhaupt erst meine Promotion ermöglicht und mich währenddessen auf die erdenklich beste und hilfreichste Art und Weise unterstützt haben. Danke, dass eure Türen stets offen waren, ich euch alle Fragen der Welt und diese auch wiederholt stellen durfte.

Weiterhin möchte ich mich bei Susanne Becker bedanken, die meine Anfrage zur Zweitgutachterin ohne Zögern akzeptiert und mich mit zahlreichen Ratschlägen ausgestattet hat.

Auch meinen Kolleginnen Sarah, Mia, Romina und Judith möchte ich danken für den konstruktiven Austausch, die therapeutisch anmutenden Gespräche während der Mittagspause sowie die zahlreichen Späße und Lacher. Auch der nicht-existente Mangel an Nervennahrung war zum größten Teil immer euch (und Nora) zu verdanken und ich war eine große Nutznießerin davon.

Ein großer Dank gilt zudem allen studentischen Hilfskräften und Master-Studentinnen, die bei der Durchführung meiner Studien geholfen haben sowie den Familien, die daran teilgenommen haben.

Auch meiner besten Freundin Louisa möchte ich danken, dass ich mich immer an dich wenden, dir mein Leid klagen und kurz danach wieder aus tiefstem Herzen mit dir lachen konnte. Zuletzt danke ich meiner Familie, vor allem meiner Mutter für die Ablenkung zu den Zeiten, in denen ich es brauchte – manchmal aber nicht wollte. Der größte Dank gilt meiner kleinen Familie, Peter und nun auch Lorenzo, für die bedingungslose Unterstützung und Zuwendung. Danke, dass ihr diese ganze Reise mit mir bestritten habt und immer für mich da wart wenn ich es brauchte.

Appendix

Original research articles

Study 1:

Ernsten, L., Körner, L. M., Schaper, M. L., Lawrenz, J., Richards, G., Heil, M., & Schaal, N. K. (2023). The association of prenatal amniotic sex hormones and digit ratio (2D:4D) in children aged 5 to 70 months: A longitudinal study. *PloS one*, *18*(3), e0282253. https://doi.org/10.1371/journal.pone.0282253

Author contributions: Ernsten, L.: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Visualization, Writing – original draft; Körner, L. M.: Conceptualization, Methodology, Writing – review & editing; Schaper, M. L.: Conceptualization, Formal analysis, Writing – review & editing; Lawrenz, J.: Conceptualization, Writing – review & editing; Richards, G.: Conceptualization, Methodology, Writing – review & editing; Heil, M.: Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – review & editing; Schaal, N. K.: Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing

Impact Factor (2022): 3.7 (Journal Citation Reports, 2023a)

PLOS ONE

RESEARCH ARTICLE

The association of prenatal amniotic sex hormones and digit ratio (2D:4D) in children aged 5 to 70 months: A longitudinal study

Luisa Ernsten[1]*, Lisa M. Körner[1,2], Marie Luisa Schaper[1], Judith Lawrenz[1], Gareth Richards[3], Martin Heil[1], Nora K. Schaal[1]

1 Department of Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany, **2** Department of General Pediatrics, Neonatology and Pediatric Cardiology, University Children's Hospital, Medical Faculty, Heinrich-Heine-University, Düsseldorf, Germany, **3** School of Psychology, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

* Luisa.Ernsten@hhu.de

Citation: Ernsten L, Körner LM, Schaper ML, Lawrenz J, Richards G, Heil M, et al. (2023) The association of prenatal amniotic sex hormones and digit ratio (2D:4D) in children aged 5 to 70 months: A longitudinal study. PLoS ONE 18(3): e0282253. https://doi.org/10.1371/journal.pone.0282253

Editor: Johannes Hönekopp, Northumbria University, UNITED KINGDOM

Received: April 6, 2022

Accepted: February 7, 2023

Published: March 23, 2023

Data Availability Statement: All data files are available from the Open Science Framework database (https://osf.io/824xp/.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The sex difference of the 2D:4D digit ratio (female > male)–a proposed marker for prenatal testosterone exposure—is well established. Studies suggest it already exists in utero and is of moderate effect size in adulthood. However, evidence for the claim that 2D:4D reflects prenatal androgen action is limited, and the sex difference may exhibit lability during childhood. In the present study, 244 mothers were recruited in the course of an amniocentesis examination (performed between gestational weeks 14 and 18). Prenatal testosterone (T) and estradiol (E) levels were determined from amniotic fluid for boys and girls. The majority (97.4%, $n = 114$) of available female T levels ($n = 117$) were found below the level of quantification. Therefore, only male amniotic fluid data ($n = 117$) could be included for the analysis of associations between amniotic sex hormones (T levels and T to E ratio (T/E)) and 2D:4D. The families were then invited to each of the five consecutive follow-ups (ages: 5, 9, 20, 40, and 70 months) where children's 2D:4D was measured for both hands. The alternative marker $D_{[r-l]}$ reflects the directional asymmetry of 2D:4D (right subtracted by left 2D:4D) and was subsequently calculated as an additional measure for prenatal T exposure. No significant correlations between amniotic T or the T/E ratio (measured between week 14 and 18 of gestation) with 2D:4D respectively $D_{[r-l]}$ were observed for any time point. There was a significant sex difference (females > males) and a significant age effect with moderate correlations of 2D:4D between time points. 2D:4D increased between 20 and 40 months and between 40 and 70 months of age. The findings raise questions regarding the applicability of 2D:4D as a marker for prenatal androgen action and are discussed in terms of the reliability of obtained digit ratio data as well as in terms of the developmental timing of amniocentesis.

Introduction

Since Manning, Scutt, Wilson, and Lewis-Jones [1] suggested the sexually differentiated ratio between the second and fourth digit length (2D:4D or digit ratio) as a marker for prenatal

testosterone (T) exposure, a vast number of studies have investigated relationships between 2D:4D and different human behaviors and traits (for reviews see [2–6]). The fascination stems from the assumption that prenatal T has organizing effects on the brain and in turn affects sex-specific behavior later in life [7]. Considering the difficulties inherent with measuring prenatal sex hormones, a convenient and easily obtainable proxy such as 2D:4D is particularly appealing. However, questions remain regarding its validity.

The sex difference in 2D:4D (female > male) is well established [8] and studies of aborted fetuses suggest that it already exists in utero [9, 10]. People characterized by atypical prenatal hormonal environments have also been reported to exhibit differences in 2D:4D. Notably, individuals with complete androgen insensitivity syndrome (CAIS; 46,XY karyotype, but female phenotype) exhibit feminized 2D:4D [11, 12]. However, the effect sizes reported for the difference between healthy controls and females with CAIS were small. Further, the variance in 2D:4D for these samples does not appear to differ from that of controls, which questions the premise that it is prenatal testosterone acting on the developing tissues that explains the effect [13]. Individuals affected by congenital adrenal hyperplasia (CAH), who are exposed to high levels of androgens during the prenatal period, show masculinized 2D:4D compared to females without CAH [14–16]. However, evidence is inconsistent and the effects sizes are relatively small [14–17]. A recent meta-analysis [16] showed only small to medium sized differences in 2D:4D between CAH patients and healthy controls. Additionally, it was noted that individual studies have typically utilized small sample sizes, and that there has been relatively little research published in this area during the past decade. Other studies have reported that men with Klinefelter syndrome (47,XXY karyotype), i.e. phenotypical males with an additional X chromosome, have a more feminized 2D:4D than unaffected controls [18, 19].

2D:4D has repeatedly been shown to correlate with various behavioral measures [20], developmental conditions [21, 22], and even various types of cancer [23, 24]. However, the validity of 2D:4D as a marker for prenatal testosterone exposure has been questioned [20, 25, 26]. Nonetheless, according to a PubMed search using the keywords "2D:4D" and "digit ratio" there were more than 100 newly published studies from 2020 till today that have investigated associations between 2D:4D and various behaviors, traits, disorders, or diseases, like muscular strength and fitness [5, 27], parental income [28], sexual preferences [29], concentration of steroids like cortisol and vitamin D [30], thyroid disease [31], and migraine in adults [32].

In order to investigate whether 2D:4D is a valid proxy for prenatal androgen levels, studies using a direct measurement of hormonal exposure are required. However, it is challenging to obtain such data in human studies due to obvious ethical considerations. Some researchers have measured prenatal sex hormone exposure from amniotic fluid obtained during the course of medically necessary amniocenteses [7, 33, 34]. The advantage of this is the timing of amniocentesis, which is normally conducted between gestational weeks 14 and 20, a timeframe in which sex differences in amniotic T and fetal serum T are highest [35–38]. Amniocentesis examinations are conducted in cases of suspected chromosomal anomalies or due to other risk factors (e.g., higher maternal age) during the second trimester of pregnancy. As this procedure is associated with a small risk of miscarriage as a consequence of invasive sampling [39, 40], non-invasive first trimester screening techniques have recently replaced many amniocentesis examinations [41]. It has therefore become less feasible to collect amniotic fluid samples, with only a few studies relating such data to subsequent phenotypic outcomes [42].

To date there are only three studies investigating associations between amniotic T and 2D:4D. Lutchmaya, Baron-Cohen, Raggatt, Knickmeyer, and Manning [43] published a prominent study showing that 2D:4D of the right hand of 2-year-old children was negatively correlated with the ratio of amniotic T to amniotic estradiol (E), referred to as the T/E ratio, measured during the second trimester of pregnancy. The sample consisted of 29 mother-infant

dyads (18 male infants). Male infants exhibited descriptively lower 2D:4D than female infants, but this sex difference was not statistically significant. No significant association emerged between 2D:4D and T or E. The authors suggested the T/E ratio has an influence on fetal growth, developmental conditions like autism, protection against early onset of breast cancer in women or myocardial infarcts in men, as well as athletic abilities in men. This study is frequently cited as evidence for 2D:4D being a valid marker for prenatal sex hormone exposure, despite its methodological limitations. Notably, the sample size was quite small and correlational analyses were not conducted separately for girls and boys, although sex was controlled for as a covariate in multiple regression analysis [43]. Furthermore, and critically, a replication of the study did not find the proposed negative correlation between prenatal T and E measured in amniotic fluid as well as the T/E ratio between weeks 15 and 20 of gestation and 2D:4D measured after 4.5 years in a sample of 66 children and their mothers [44]. Richards Browne, and Constantinescu [44] found, in contrast to the assumed correlations based on Lutchmaya et al. [43], no associations between T as well as the T/E ratio and 2D:4D for either girls or boys. The authors also examined the difference between right and left 2D:4D ($D_{[r-l]}$). This variable is discussed as an additional measure, with lower values thought to indicate a higher intrauterine androgen exposure [45–47]. However, this variable also showed no association with prenatal androgens [44]. The third amniocentesis study from Ventura, Gomes, Pita, Neto, and Taylor found the expected negative relationship between amniotic T and 2D:4D measured a few days after birth [48]. However, this was present only for the left hand and only in newborn girls (significant correlation for left 2D:4D, statistical trend for right 2D:4D [48]).

Another issue in 2D:4D research is limited evidence of temporal stability, especially in young cohorts. Based on the assumption that 2D:4D is (at least partially) determined by prenatal androgens one can assume that the temporal stability should be high, especially before the onset of puberty. However, a review of relevant studies revealed limited evidence for the temporal stability of 2D:4D, and, moreover, the sex difference in prepubertal cohorts [49]. Longitudinal studies have shown an increase of 2D:4D with age, although with moderate to high correlations between the time points [20, 50–52]. Across these studies, the sex difference in 2D:4D was in general small and did not always reach statistical significance, although the direction (female > male) was always the same [50, 51]. A longitudinal study of 0-2-year-olds (age: 2 weeks, 12 months, 24 months) reported a significant sex effect for 2D:4D measured two weeks after birth as well as a decrease in 2D:4D in the first year and an increase in the second year of life [53]. Interestingly, at 12 and 24 months there were no significant sex differences in 2D:4D. However, cross-sectional studies revealed a sex difference as well as an increase in 2D:4D for children between the ages of 2 and 5 years [54] and in a cohort aged between 5 and 17 years [55].

Manning and Fink [56] have promoted the assumption that 2D:4D remains independent of skeletal growth, with an earlier cross-sectional study [1] reporting a stable sex difference and no significant age effects between cohorts of 2-25-year-old subjects. However, the majority of longitudinal as well as cross-sectional studies do not indicate the temporal stability of 2D:4D in younger cohorts (especially under the age of two), which may be affected by skeletal and overall growth. Studies with older cohorts reveal greater and more stable sex differences (specifically adult and adolescent samples [8]), in a timeframe where the finger lengths remain more or less the same [57].

It becomes apparent that the available literature does not allow precise conclusions on the hypothesized validity and reliability of 2D:4D as a marker of prenatal androgen action. Nonetheless, 2D:4D remains a popular tool used by researchers attempting to investigate the effects of prenatal androgen action. For these reasons, the current study examines amniocentesis data of pregnant women and their offspring's 2D:4D in a longitudinal design. This approach

provides an opportunity to test whether 2D:4D is a valid proxy for prenatal T levels, and further to test its temporal stability within a young cohort. Amniotic fluid was obtained between gestational weeks 14 and 18 in a sample of 244 pregnant women in Germany. The overall levels of amniotic T and E were determined, and the T/E ratio was calculated for further analysis. Unfortunately, data were only available for the male subsample as female T levels were under the limit of detection or quantification. Children's 2D:4D on both hands was measured at 5, 9, 20, 40, and 70 months of age. With respect to the hypothesis that 2D:4D acts as a marker for prenatal androgen exposure, negative correlations between amniotic T as well as the T/E ratio and 2D:4D were predicted for the male subsample at each time point. As other studies also analyzed a possible association between amniotic T and T/E ratio with $D_{[r-l]}$, this additional variable was included. We also predicted that 2D:4D would show moderate to high correlations between the measurement time points, as well as a stable sex difference (females > males for all time points). Further, based on the available literature, an increase of 2D:4D with age was expected.

Methods

Participants

Two-hundred forty-four mothers who underwent amniocentesis at a practice of gynecologists and human genetics (*Praenatal.de*) in Düsseldorf, Germany, were recruited between 2010 and 2012. The children were born between January 2011 and February 2013 and were invited to the Department of Experimental Psychology at the University of Düsseldorf, Germany, on five occasions. More specifically, participants took part in follow-up research at the ages of 5 months (T1), 9 months (T2), 20 months (T3), 40 months (T4), and 70 months (T5) as part of another study that also obtained behavioral data published elsewhere [58, 59]. Although every initially recruited family was invited to participate, not every family took part in all consecutive measurement time points resulting in different sample sizes (for mean ages, sample sizes and sex distributions, see Table 1). Mothers were between 22 and 48 years old (M = 38.13 years, SD = 3.50 years) when they gave birth. All families were white. Of the 244 initially recruited mothers (n = 123 male fetuses and n = 121 female fetuses), n = 231 were used for data analysis, as for n = 2

Table 1. Number and age (in months) of participants at T1-T5.

Time Point		Boys	Girls	Total
T1: 5 months	*N*	114	111	225
	M	5.40	5.45	5.43
	SD	0.29	0.31	0.30
T2: 9 months	*N*	101	91	192
	M	9.38	9.36	9.37
	SD	0.35	0.39	0.37
T3: 20 months	*N*	86	80	166
	M	20.54	20.52	20.53
	SD	0.34	0.44	0.39
T4: 40 months	*N*	80	78	158
	M	40.51	40.56	40.53
	SD	0.49	0.65	0.57
T5: 70 months	*N*	73	74	147
	M	70.69	70.68	70.69
	SD	0.92	1.38	1.17

https://doi.org/10.1371/journal.pone.0282253.t001

female samples there was no hormonal data as well as no hand scan for any of the measurement time points, and for $n = 5$ male and $n = 6$ female samples there were no corresponding hand scans for any of the measurement time points, leaving $n = 118$ male samples and $n = 113$ female samples. Of the 244 amniocentesis examinations, $n = 1$ male and $n = 4$ female amniotic samples were missing, however, as hand scans were available, these were used for data analysis. Further, the majority of female fetuses T values (97.4%) were below the limit of detection (0.02 ng/ml) or between the limit of detection and the limit of quantification (0.05 ng/ml) and therefore, the female sample was not included for the respective data analysis including hormonal data.

For recruitment, mothers gave their written informed consent for the use of the data obtained through amniocentesis as well as for re-contacting for the following measurements. At every following measurement (T1-T5) the parents again gave written informed consent for participation. The longitudinal study was approved by the local Ethics Committee of the Science Faculty of the University of Düsseldorf, Germany.

Materials

Prenatal hormone concentration measurements. Total prenatal T and E levels (in ng/ml) were determined in amniotic fluid from amniocentesis samples of mothers recruited from *Praenatal.de* (Düsseldorf, Germany). The amniocentesis examinations were carried out between weeks 14 and 18 of gestation ($M = 14.78$, $SD = 0.84$). Amniotic fluid samples were assayed with ultra-performance liquid chromatography and tandem mass spectrometry (described elsewhere [58, 59]). For data analysis, $n = 117$ male samples could be included.

As discussed by Lutchmaya et al. [43], the ratio between amniotic T and E (amniotic T/E) may have a similar impact on early fetal and postnatal development as amniotic T alone. Therefore, we included this ratio in our analysis. Amniotic T/E was calculated as follows:

$$Amniotic\ T/E = \frac{amniotic\ T\ levels\ \left[\frac{ng}{ml}\right]}{amniotic\ E\ levels\ \left[\frac{ng}{ml}\right]}$$

2D:4D. Both hands of the children were scanned using a FIJUTSI fi-60F image scanner at every time point (T1-T5), as hand scans yield both larger sex differences [8] and a higher measurement precision [60, 61] compared to direct measurements. The freeware program *Autometric* [62] was used to measure the ratio between the second and the fourth digit length (2D:4D). Each digit was measured as the length of the midpoint of the ventral proximal crease to the tip in *pixels* using a 100-dpi monitor (100 *pixels* = 2.54 *cm*). Scans in which the fingertips or the ventral creases were not distinct were excluded as well as scans that could not be used for calculation of 2D:4D (e.g. only the length of the second finger was measurable but not of the fourth finger). Two raters, each blind to the sex of the children, measured all hand scans (max. 10 per child). Intraclass correlations indicated high inter-rater reliabilities (all ICCs $> .90$). The measurements of the two raters were averaged to increase reliability. 2D:4D was calculated as follows:

$$2D:4D = \frac{length\ of\ the\ second\ digit\ [pixels]}{length\ of\ the\ fourth\ digit\ [pixels]}$$

$D_{[r\text{-}l]}$.

Additionally, the measure $D_{[r\text{-}l]}$ defined as the difference between right and left 2D:4D was computed:

$$D_{[r-l]} = right\ 2D:4D - left\ 2D:4D$$

Statistical analyses

A multilevel linear regression model [63, 64] using the R packages lme4 and lmerTest [65–67] with participants as random effects was used to examine the association between amniotic T as well as amniotic T/E and 2D:4D in boys. Age (months since birth, centralized to the participant mean), amniotic T level (logarithmized to achieve normal distribution, centralized to the grand mean) and week of pregnancy the amniocentesis took place (centralized to the grand mean) were entered as predictors of 2D:4D in the first model; age, amniotic T/E level (logarithmized to achieve normal distribution, centralized to the grand mean) and week of pregnancy the amniocentesis took place were entered as predictors in the second model.

The same multilevel linear regression models [63, 64] with participants as random effects were used to examine the association between amniotic T as well as amniotic T/E and $D_{[r-l]}$ in boys. Multilevel linear regression models with participants as random effects were also used to test whether age, sex, and hand were associated with 2D:4D. The advantage of this analysis is that it accounts for participant heterogeneity as well as missing data (distribution of participants for each measurement time point can be seen in Table 1). Age (months since birth, centralized to the participant mean), sex (first coded as boys = 0, girls = 1, then centralized to the grand mean), and hand (first coded as left = 0, right = 1, then centralized to the participant mean) were entered as predictors of 2D:4D. Knickmeyer, Woolson, Hamer, Konneker, and Gilmore [53] also used mixed models in their longitudinal study of 2D:4D. To follow-up on the association between age and 2D:4D, four additional multilevel linear regression models comparing the influence of age between T1 and T2, T2 and T3, T3 and T4, and T4 and T5 were calculated. Lastly, Pearson correlations of 2D:4D between the different time points were conducted and interpreted as follows: small correlation $r \geq .10$, moderate correlation $r \geq .30$, and large correlation $r \geq .50$ [68].

Results

2D:4D and amniotic T and T/E

Boys had mean T levels of $M = 0.099$ ng/ml ($SD = 0.065$ ng/ml, range: 0.014–0.339 ng/ml, see Fig 1 for distribution) and mean E levels of $M = 0.146$ ng/ml ($SD = 0.080$ ng/ml, range: 0.047–0.496 ng/ml). There were no significant associations between amniotic T or amniotic T/E and 2D:4D, and no interactions with any other factor in boys (controlling for age at measurement of 2D:4D and week of pregnancy of amniocentesis) as indicated by two multilevel linear regression models (see Tables 2 and 3).

$D_{[r-l]}$ and amniotic T and T/E

There were no significant associations between amniotic T or amniotic T/E and $D_{[r-l]}$, and no interactions with any other factor in boys (controlling for age at measurement of $D_{[r-l]}$ and week of pregnancy of amniocentesis) as indicated by two multilevel linear regression models (see Tables 4 and 5).

Temporal stability of 2D:4D

Girls had larger 2D:4D compared to boys at every measurement time point. On a descriptive level, boys' 2D:4D decreased from T1 to T2 and increased from T2 to T3, T3 to T4 and T4 to T5; in girls, right 2D:4D increased whereas left 2D:4D decreased from T1 to T2, 2D:4D of both hands decreased from T2 to T3, and showed an increase from T3 to T4 and T4 to T5 (see Table 6 and Fig 2). The multilevel linear regression model with age, sex, and hand as predictors of 2D:4D (see Table 7) showed a significant main effect of sex. Moreover, there was a

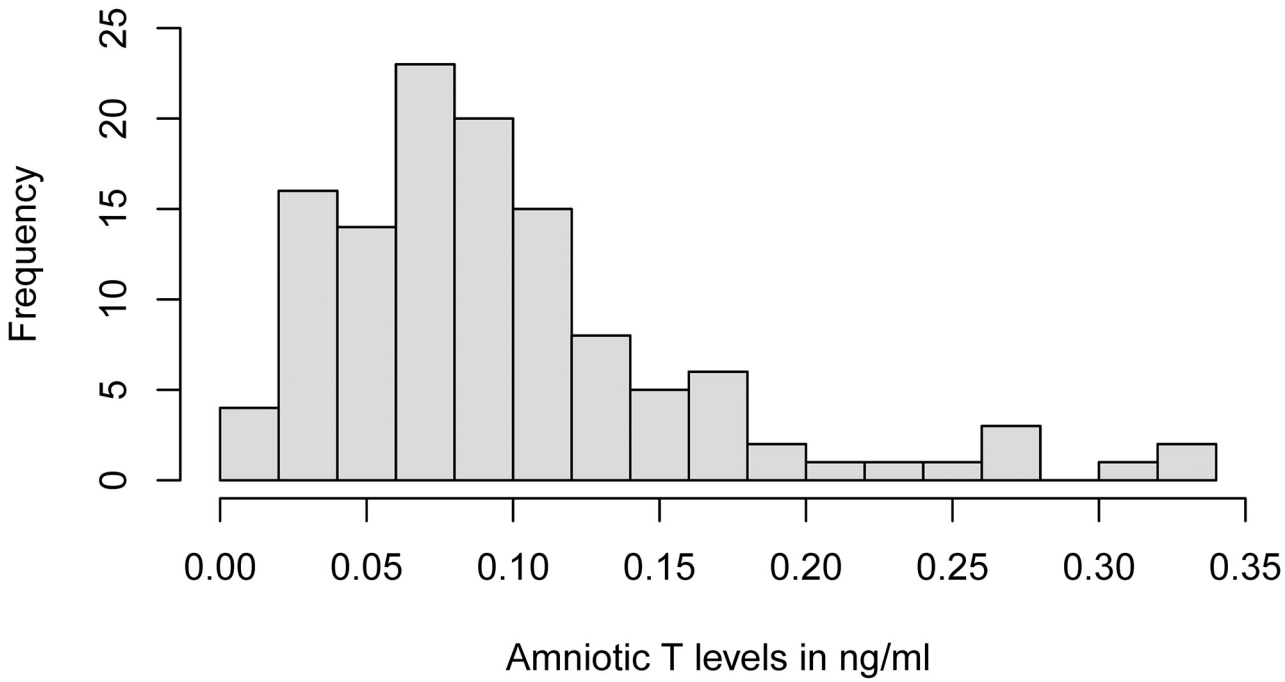


Fig 1. Distribution of male amniotic T-levels (n = 117).

https://doi.org/10.1371/journal.pone.0282253.g001

significant main effect of age indicating that 2D:4D increased with age. The interaction between sex and age was marginally significant (p = .050) and the three-way interaction between sex, age, and hand was significant (p = .036). Follow-up analyses on these interactions revealed that a) the effects of age and sex were significant in both hands (all $p < .001$), b) in the right hand, the effect of age was stronger in boys compared to girls (p = .006), and c) in the left hand, the effect of age was of equal size for boys and girls (p = .761). There was no main effect of hand, no interaction between sex and hand, and no interaction between age and hand. The additional multilevel linear regression models to follow up on the effect of age showed that age

Table 2. Multilevel linear model with age, amniotic T and week of pregnancy as predictors, and boys' 2D:4D as outcome.

Effect	Estimate	*SE*	*df*	*t*	*p*
Intercept	0.9329	0.0023	103.3	407.65	< .001
Age***	0.0004	0.0000	500.4	9.34	< .001
Amniotic T	0.0003	0.0036	101.4	0.08	.939
Week of Pregnancy	0.0043	0.0028	103.4	1.54	.127
Age x Amniotic T	0.0001	0.0001	500.4	0.94	.346
Age x Week of Pregnancy	-0.0000	0.0001	500.4	-0.80	.422
Amniotic T x Week of Pregnancy	0.0047	0.0049	129.9	0.96	.339
Age x Amniotic T x Week of Pregnancy	0.0002	0.0001	500.4	1.26	.207

Note. Estimates represent unstandardized regression weights. Analyses were performed with the R procedures lme4 and lmerTest with restricted maximum likelihood estimation. Testosterone = logarithmized testosterone level from amniotic fluid; Week of Pregnancy = week of pregnancy of amniocentesis examination; *SE* = Standard error.

***$p < .001$

https://doi.org/10.1371/journal.pone.0282253.t002

Table 3. Multilevel linear model with age, amniotic T/E and week of pregnancy as predictors, and boys' 2D:4D as outcome.

Effect	Estimate	*SE*	*df*	*t*	*p*
Intercept	0.9329	0.0023	103.3	408.41	< .001
Age***	0.0004	0.0000	500.3	9.32	< .001
Amniotic T/E	-0.0014	0.0029	101.7	-0.50	.616
Week of Pregnancy	0.0041	0.0028	104.2	1.45	.149
Age x Amniotic T/E	0.0000	0.0001	500.3	0.56	.573
Age x Week of Pregnancy	-0.0001	0.0001	500.3	-0.94	.349
Amniotic T/E x Week of Pregnancy	0.0034	0.0035	109.7	0.96	.341
Age x Amniotic T/E x Week of Pregnancy	0.0001	0.0001	500.3	1.38	.169

Note. Estimates represent unstandardized regression weights. Analyses were performed with the R procedures lme4 and lmerTest with restricted maximum likelihood estimation. *SE* = Standard error.

***$p < .001$

https://doi.org/10.1371/journal.pone.0282253.t003

Table 4. Multilevel linear model with age, amniotic T and week of pregnancy as predictors, and boys' $D_{[r-l]}$ as outcome.

Effect	Estimate	*SE*	*df*	*t*	*p*
Intercept	0.0002	0.0028	102.0	0.06	.956
Age†	0.0002	0.0001	157.8	1.95	.053
Amniotic T	0.0038	0.0043	88.9	0.87	.388
Week of Pregnancy	-0.0015	0.0034	104.9	-0.43	.671
Age x Amniotic T	-0.0001	0.0001	157.8	-0.58	.564
Age x Week of Pregnancy	-0.0002	0.0001	157.8	-1.29	.198
Amniotic T x Week of Pregnancy	0.0069	0.0067	101.4	1.03	.305
Age x Amniotic T x Week of Pregnancy	-0.0003	0.0002	157.8	-1.42	.159

Note. Estimates represent unstandardized regression weights. Analyses were performed with the R procedures lme4 and lmerTest with restricted maximum likelihood estimation. Testosterone = logarithmized testosterone level from amniotic fluid; Week of Pregnancy = week of pregnancy of amniocentesis examination; *SE* = Standard error.

†$p < .10$

https://doi.org/10.1371/journal.pone.0282253.t004

Table 5. Multilevel linear model with age, amniotic T/E and week of pregnancy as predictors, and boys' $D_{[r-l]}$ as outcome.

Effect	Estimate	*SE*	*df*	*t*	*p*
Intercept	0.0003	0.0028	100.9	0.10	.923
Age†	0.0002	0.0001	156.9	1.94	.054
Amniotic T/E	-0.0036	0.0034	95.5	-1.03	.304
Week of Pregnancy	-0.0011	0.0034	103.0	-0.31	.759
Age x Amniotic T/E	0.0000	0.0001	156.9	0.24	.810
Age x Week of Pregnancy	-0.0001	0.0001	156.9	-1.18	.240
Amniotic T/E x Week of Pregnancy	0.0034	0.0045	107.8	0.75	.453
Age x Amniotic T/E x Week of Pregnancy	-0.0001	0.0002	156.9	-0.74	.463

Note. Estimates represent unstandardized regression weights. Analyses were performed with the R procedures lme4 and lmerTest with restricted maximum likelihood estimation. *SE* = Standard error.

†$p < .10$

https://doi.org/10.1371/journal.pone.0282253.t005

Table 6. Sample sizes, means and standard deviations for right and left 2D:4D at T1-T5.

		Right hand			Left hand		
		N	*M*	*SD*	*N*	*M*	*SD*
T1	Boys	71	.924	.039	70	.926	.036
	Girls	65	.946	.036	85	.941	.033
	Overall	136	.935	.039	155	.934	.035
T2	Boys	56	.916	.037	68	.923	.038
	Girls	49	.947	.034	60	.938	.037
	Overall	105	.931	.039	128	.930	.038
T3	Boys	51	.923	.039	50	.926	.031
	Girls	51	.939	.032	50	.934	.033
	Overall	102	.931	.036	100	.930	.032
T4	Boys	56	.937	.028	60	.934	.033
	Girls	65	.955	.030	67	.954	.034
	Overall	121	.947	.030	127	.945	.035
T5	Boys	73	.953	.024	72	.948	.030
	Girls	74	.962	.033	74	.958	.031
	Overall	147	.958	.029	146	.953	.031

Note. Age at T1: 5 months, at T2: 9 months, at T3: 20 months, at T4: 40 months, and at T5: 70 months.

https://doi.org/10.1371/journal.pone.0282253.t006

was marginally associated with 2D:4D between T1 and T2, but not between T2 and T3. There were significant age effects between T3 and T4, as well as between T4 and T5 (see Table 8). Table 9 indicates that the sex effect remained stable for each measurement time point separately. There was no main effect for hand and no significant interaction between hand and sex for each measurement point separately (see Table 9).

Regarding the reliability of 2D:4D between measurement time points, Pearson correlations revealed significant positive correlations for the right hand between T1 and T2, $r = .42$, $p = .001$, between T2 and T3, $r = .29$, $p = .037$, between T3 and T4, $r = .50$, $p < .001$, and between T4 and T5, $r = .66$, $p < .001$. For the left hand, significant positive correlations could be found between T1 and T2, $r = .47$, $p < .001$, between T2 and T3, $r = .30$, $p = .023$, between T3 and T4, $r = .39$, $p = .001$, and between T4 and T5, $r = .61$, $p < .001$.

Discussion

The current study used a longitudinal design and had two main aims: (1) to determine whether 2D:4D is associated with prenatal androgen (T and T/E ratio) concentrations measured from amniotic fluid obtained between 14 and 18 weeks of gestation, and (2) to examine the stability of sex differences in 2D:4D at 5, 9, 20, 40, and 70 months of age. There was no significant correlation between 2D:4D respectively $D_{[r-l]}$ and either amniotic T or the T/E ratio in boys for any time point. For girls, amniotic T and T/E ratio could not be determined and therefore no analysis was performed. At each time point, a consistent sex effect was revealed with girls showing larger 2D:4D ratios than boys. Age was associated with 2D:4D, with post-hoc-analyses revealing a marginally significant decrease in 2D:4D from 5 to 9 months of age and statistically significant increases in 2D:4D from 20 to 40 and from 40 to 70 months of age. Significant positive correlations for 2D:4D between different time points suggest a moderate level of stability during infancy and childhood. However, when taken together, the results of this study indicate that 2D:4D exhibits considerable lability during infancy and also question the validity of 2D:4D as a marker of prenatal androgen exposure.

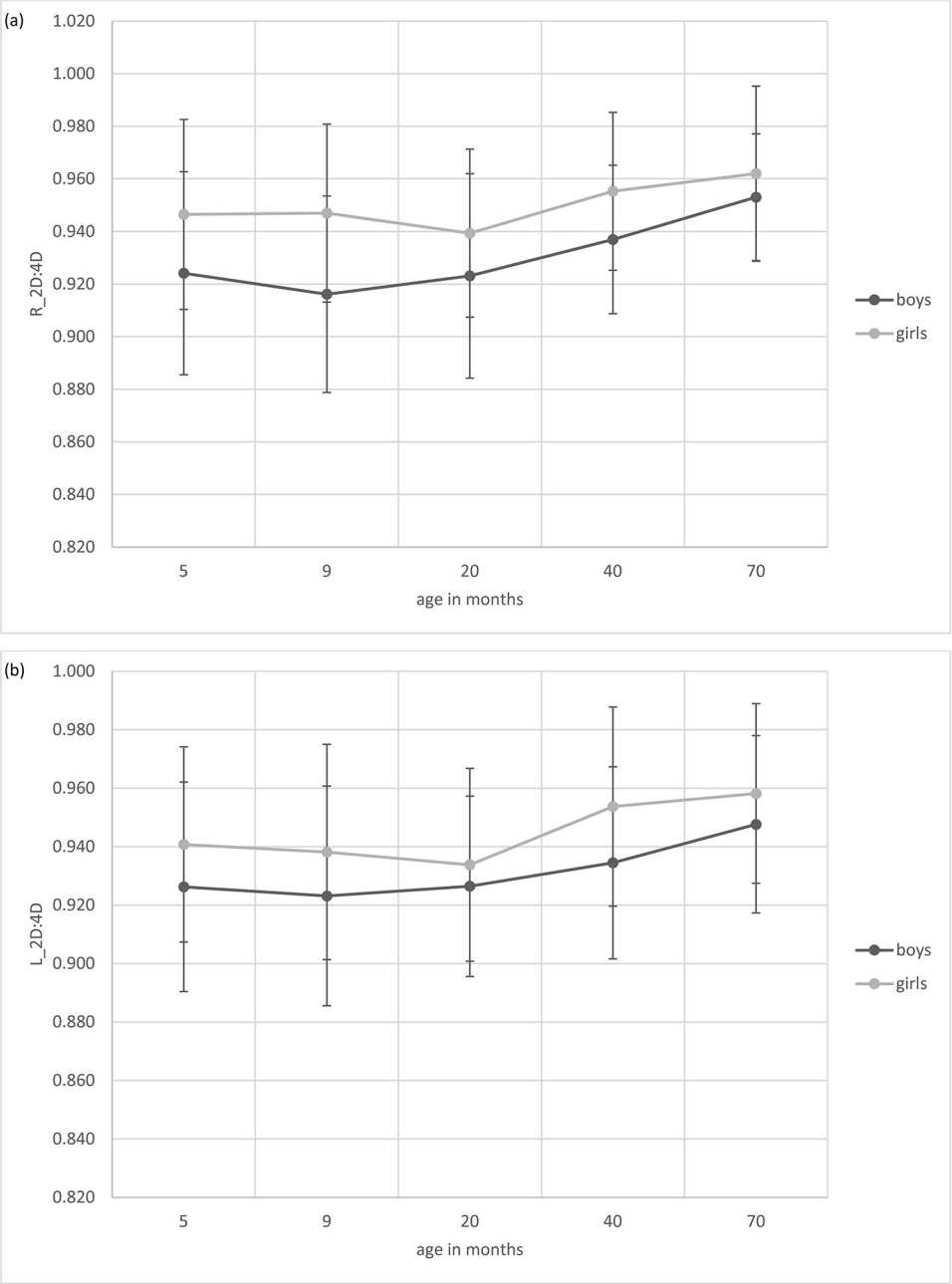


Fig 2. Means and Standard Deviations of (a) Right and (b) Left 2D:4D for every Time Point in Girls and Boys.

https://doi.org/10.1371/journal.pone.0282253.g002

The current state of research does not allow precise conclusions about the hypothesis that 2D:4D is a marker of prenatal androgen action, with heterogeneity in study designs and differing results providing an inconsistent picture. In our study, we did not find any associations between 2D:4D respectively $D_{[r\text{-}l]}$ and either amniotic T or T/E in boys for any of the postnatal time points. However, the absence of female amniocentesis data is especially unfavorable. This is because it precluded the opportunity to try to replicate the finding of Ventura et al. [48] of a significant negative association between T levels and left 2D:4D in girls (but not boys). Ventura

Table 7. Multilevel linear model with sex, age and hand as predictors, and 2D:4D as outcome.

Effect	Estimate	*SE*	*df*	*t*	*p*
Intercept	0.9402	0.0016	212.0	577.57	< .001
Sex***	0.0138	0.0033	212.0	4.24	< .001
Age***	0.0004	0.0000	1025.0	11.72	< .001
Hand	0.0017	0.0015	1025.0	1.08	.280
Sex × Age†	-0.0001	0.0001	1025.0	-1.96	.050
Sex × Hand	0.0041	0.0031	1025.0	1.33	.185
Age × Hand	0.0001	0.0001	1039.0	1.01	.311
Sex × Age × Hand*	-0.0003	0.0001	1039.0	-2.10	.036

Note. Estimates represent unstandardized regression weights. Analyses were performed with the R procedures lme4 and lmerTest with restricted maximum likelihood estimation. For sex, 0 = boys, 1 = girls. For hand, 0 = left, 1 = right. *SE* = Standard error.

†$p < .10$,

*$p < .05$,

***$p < .001$

https://doi.org/10.1371/journal.pone.0282253.t007

et al. [48] hypothesized that because boys are already exposed to higher prenatal T levels than girls, increased T above the average might not have a similar impact on male compared to female 2D:4D, and so may potentially account for the non-significant correlation in boys. On the other hand, Lutchmaya et al. [43] found a statistically significant negative association between the T/E ratio and 2D:4D that remained even when controlling for sex. The attempted replication of Lutchmaya et al. [43] by Richards et al. [44] did not find any of the hypothesized correlations for either sex. This was despite the higher statistical power associated with there being more than twice as many mother-child-dyads as in the original study. These differences in findings may stem from a bias concerning the reliability of data obtained through amniocentesis. Notably, amniocentesis samples reflect only a small time window during prenatal development. Lutchmaya et al. [43] reported that the examinations took place during the second trimester of pregnancy without specifying the mean and range of gestational weeks. Other studies in this area examined amniocentesis samples collected between 15 and 22 weeks [44],

Table 8. Multilevel linear model with age as predictor and 2D:4D as outcome.

Measurement		Estimate	*SE*	*df*	*t*	*p*
T1 vs. T2†	Intercept	0.9331	0.0021	219.5	447.44	< .001
	Age	-0.0013	0.0008	326.5	-1.71	.089
T2 vs. T3	Intercept	0.9307	0.0021	177.8	445.85	< .001
	Age	0.0002	0.0003	253.5	0.45	.655
T3 vs. T4***	Intercept	0.9392	0.0021	171.7	455.06	< .001
	Age	0.0008	0.0001	279.7	5.36	< .001
T4 vs. T5***	Intercept	0.9504	0.0020	167.6	484.25	< .001
	Age	0.0004	0.0001	373.7	5.35	< .001

Note. Estimates represent unstandardized regression weights. Analyses were performed with the R procedures lme4 and lmerTest with restricted maximum likelihood estimation. For sex, 0 = boys, 1 = girls. For hand, 0 = left, 1 = right. T1 = 5 months (4.63–6.87 months), T2 = 9 months (7.59–10.25 months), T3 = 20 months (18.56–21.65 months), T4 = 40 months (38.28–42.64 months), T5 = 70 months (67.16–74.52 months). *SE* = Standard error.

†$p < .10$,

***$p < .001$

https://doi.org/10.1371/journal.pone.0282253.t008

Table 9. Multilevel linear model with hand and sex as predictors and 2D:4D as outcome for each measurement time point.

Measurement		Estimate	*SE*	*df*	*t*	*p*
T1	Intercept	0.9348	0.0024	181.9	388.51	< .001
	Hand	0.0052	0.0035	120.2	1.50	.135
	Sex***	0.0188	0.0048	182.0	3.90	< .001
	Hand x Sex	0.0024	0.0069	120.2	0.35	.731
T2	Intercept	0.9303	0.0027	140.4	346.95	< .001
	Hand	-0.0019	0.0045	84.6	-0.43	.671
	Sex***	0.0213	0.0054	140.4	3.96	< .001
	Hand x Sex	0.0148	0.0089	84.6	1.65	.102
T3	Intercept	0.9307	0.0025	119.2	379.46	< .001
	Hand	-0.0013	0.0053	81.6	-0.24	.809
	Sex*	0.0115	0.0049	119.2	2.35	.021
	Hand x Sex	0.0170	0.0106	81.6	1.60	.113
T4	Intercept	0.9456	0.0023	139.1	408.97	< .001
	Hand	0.0025	0.0033	108.4	0.76	.451
	Sex***	0.0173	0.0046	139.8	3.74	< .001
	Hand x Sex	-0.0044	0.0066	108.4	-0.67	.507
T5	Intercept	0.9554	0.0021	144.1	462.26	< .001
	Hand	0.0042	0.0027	143.3	1.57	.119
	Sex*	0.0096	0.0041	144.1	2.32	.022
	Hand x Sex	-0.0009	0.0054	143.3	-0.17	.864

Note. Estimates represent unstandardized regression weights. Analyses were performed with the R procedures lme4 and lmerTest with restricted maximum likelihood estimation. For sex, 0 = boys, 1 = girls. For hand, 0 = left, 1 = right. T1 = 5 months (4.63–5.43 months), T2 = 9 months (7.59–10.25 months), T3 = 20 months (18.56–21.65 months), T4 = 40 months (38.28–42.64 months), T5 = 70 months (67.16–74.52 months). *SE* = Standard error.

*p < .05,

***$p < .001$

https://doi.org/10.1371/journal.pone.0282253.t009

16 and 24 weeks (M = 17.2 [48]), and 14 and 18 weeks (M = 14.78; current study). This may account for the differing results, as literature suggests considerable fluctuations of T during the second trimester as well as a peak around week 17 [35, 36]. A more recent study of Kuijper, Ket, Caanen, and Lambalk [69] reports that T levels seem to increase between weeks 14 and 22, with no clear peak for boys and a peak around week 18 for girls. Due to risks and ethical considerations, repeated sampling of amniotic fluid in humans is not possible. Nevertheless, amniocentesis data may allow the most precise information available to assess the hormonal intrauterine environment of a developing human fetus [33]. As Hollier, Keelan, Hickey, Maybery, and Whitehouse [70] discussed, to date there is no "gold standard" to examine hormone levels to which the fetus is exposed during pregnancy. Other studies have revealed inconsistent and mostly non-significant associations between children's 2D:4D and alternative measures like maternal plasma or cord blood samples to assess T levels in pregnant women [71–74].

Another point to consider is that prenatal androgens are unlikely to be the only determining factor for 2D:4D as the sex difference in prenatal androgen levels is much larger than the sex difference observed for 2D:4D [11, 75]. Furthermore, studies of individuals with CAIS show differences in 2D:4D despite T having no physiologic effect [11, 12]. However, Zheng and Cohn [76] showed that active signaling and density of androgen and estrogen receptors differ in the second and fourth digit, which may lead to the sexually differentiated pattern of 2D:4D (female > male) in a mouse model. The sex difference in 2D:4D in mice was present as

early as the 17th day of embryonic life [76]. A direct manipulation of the intrauterine environment of mice showed that an activation or blocking of the androgen receptor resulted in changes of the total paw and digit length as well as 2D:4D [77]. That is, an activation of the androgen receptor led to an increase of 2D:4D in males, whereas a blocking of the androgen receptor led to a decrease of 2D:4D in females [77]. This finding is also in line with Swift-Gallant, Di Rita, Coome, and Monks [78] who found an increase in 2D:4D in mice with global androgen receptor overexpression compared to wildtype mice and mice with neural-specific androgen receptor overexpression regardless of sex. Other studies administered T directly as an intramuscular injection into the dam and examined 2D:4D in the offspring. Talarovičová, Kršková, and Blažeková [79] found a significant effect of T administration on the length of the second and fourth digit as well as on 2D:4D. Administration of T, compared to sesame oil as a control, led to a lengthening of the fourth digit, a shortening of the second digit, and a smaller 2D:4D. Contrary, another study found shortened second and fourth digits in the offspring after T (respectively T combined with flutamide) administration compared to a control group that was supplemented with olive oil only, however, no effect of T administration on 2D:4D [80]. These findings indeed suggest an influence of androgens on digit lengths and 2D:4D but also reveal the complex interplay of hormones, chromosomes, and gonads to determine the sex and sex-specific characteristics of organisms and the influence on digit growth.

Despite the current study revealing stable sex differences with girls exhibiting larger 2D:4D compared to boys, 2D:4D appears unable to reliably discriminate between the sexes as the magnitude of 2D:4D in girls and boys reveals a large overlap. This could be shown in various other studies reporting the expected pattern of 2D:4D, however, without significant sex differences and/or only small effect sizes [50–53, 81]. Furthermore, girls show considerable variability in 2D:4D yet relative homogeneity in amniotic T levels [36, 44]; this was also true for the current study, as 97.4% of girls' T levels were under the limit of detection or quantification. Another observation that questions the validity of 2D:4D is its association with age in young cohorts, an effect reported not only in the current study but also in various others within the literature [51–55, 81, 82]. The significant effect of age indicates that 2D:4D continues to develop during childhood. Importantly, the reliability of 2D:4D at this time is also questionable, as the current study observed only moderately sized correlations between time points. Likewise, the association between age and 2D:4D makes the results of Lutchmaya et al. [43] and Richards et al. [44] difficult to compare, as 2D:4D was measured at 2 and 4.5 years of age respectively.

Notably, research does not support reliable sex differences of 2D:4D in younger cohorts, as noted in another study of our group [49], and it appears that sex differences become larger and more stable during adolescence and adulthood [8, 20, 50–52]. Prepubertal girls and boys show nearly the same concentrations of fluctuating sex hormones [83]. There is only a short period of time during the first 6 months after birth in which specifically T levels increase in boys, whereas the levels do not differ to a greater extent between girls and boys until the onset of puberty [84, 85]. A comparison of characteristics that can be associated with the influence of prenatal sex hormones should be highly reliable between these two time points—birth and puberty. However, studies failed to replicate a stable sex difference in young cohorts [56, 86], questioning the assumption that prenatal amniotic T levels explain a significant proportion of variance in 2D:4D. The measurement of digit lengths in young cohorts compared to adult cohorts may face other methodological issues that challenge the reliability and validity of the measurement itself. Young infants show a much higher proportion of body fat tissue compared to adults [87], which may specifically impair indirect measurement techniques like hand scans, as hand creases tend to shift more in between different measurements when pressed onto the scanner glass. Furthermore, very young children and infants may be less likely to cooperate in

the measurement procedure like adults normally do. To control for this, another more direct measurement technique may be a better approach to examine 2D:4D in young cohorts and, compared to indirect measurement techniques, would also yield greater effect sizes [88]. In fact, McIntyre, Ellison, Lieberman, Demerath, and Towne [51] and McIntyre, Cohn, and Ellison [52] used radiographs in their cohorts aged 1 month up to 18 years. However, they found only low to moderate differences in 2D:4D between males and females with more stable differences emerging with age. The lack of comparability between different studies in terms of measurement time point and measurement technique impairs the interpretation of different results specifically in prepubertal cohorts.

Lastly, there are some limitation of the current study that should be addressed. It has already been discussed that the timing of the amniocentesis examination may account for different results between studies. Further, there is a relative large interval between human limb development and amniocentesis examinations as the human limb development starts much earlier, approximately during week 4 of gestation and by the end of week 8 of gestation the upper and lower limbs are already mostly developed [89, 90], compared to amniocenteses which are usually performed around week 15^{+0} of gestation [91]. Therefore, prenatal hormone levels obtained through amniotic sampling may not fully reflect the intrauterine environment during limb development. Additionally, amniotic fluid only reflects the hormonal environment the fetus is subjected to, not the actual hormonal concentrations of the fetus [69]. Nevertheless, amniotic fluid sampling may be the best method available to obtain information about fetal hormonal environment, due to obvious ethical considerations. Also the 2D:4D digit ratio is facing important influencing factors like the number and order of siblings and their sex (however, with contradictory results, see Králík, Hupková, Zeman, Hložek, Hlaváček, Slováčková, et al. [92] and Saino, Leoni, and Romano [93]). Further, as it is hypothesized that 2D:4D differs between different ethnic groups [94, 95], the current results are not generalizable as they only refer to the German and solely white population.

The motivation to obtain reliable and valid information about the intrauterine hormonal environment by simply measuring digit lengths is understandable considering the difficulties associated with obtaining more direct measurements. The literature does provide some evidence for 2D:4D as it exhibits a moderate sized sex difference as well as associations with other sexually differentiated characteristics. However, the evidence that 2D:4D reflects prenatal androgen action is more limited and a comparison of available studies is difficult. Therefore, 2D:4D does not seem to be a feasible proxy for intrauterine androgen exposure. Although the sex difference in 2D:4D shows a good replicability, especially in adult samples, the hypothetical fundament of its origins is not sufficiently examined. If 2D:4D shall be used as a marker for prenatal androgen action it is necessary to examine the developmental basis. Further, instead of conducting novel research on associations between 2D:4D and various other characteristics, research should focus on replications with highly comparable study designs and on testing the underlying hypothesis that 2D:4D is a valid marker of intrauterine androgen action.

Supporting information

S1 Fig. Data inclusion and exclusion chart.
(TIF)

Author Contributions

Conceptualization: Luisa Ernsten, Lisa M. Körner, Marie Luisa Schaper, Judith Lawrenz, Gareth Richards, Martin Heil, Nora K. Schaal.

Data curation: Luisa Ernsten.

Formal analysis: Luisa Ernsten, Marie Luisa Schaper, Nora K. Schaal.

Investigation: Martin Heil.

Methodology: Luisa Ernsten, Lisa M. Körner, Gareth Richards, Martin Heil, Nora K. Schaal.

Project administration: Luisa Ernsten.

Resources: Martin Heil.

Supervision: Martin Heil, Nora K. Schaal.

Visualization: Luisa Ernsten.

Writing – original draft: Luisa Ernsten.

Writing – review & editing: Lisa M. Körner, Marie Luisa Schaper, Judith Lawrenz, Gareth Richards, Martin Heil, Nora K. Schaal.

References

1. Manning J, Scutt D, Wilson J, Lewis-Jones D. The Ratio of 2nd to 4th Digit Length: A Predictor of Sperm Numbers and Concentrations of Testosterone, Luteinizing Hormone and Oestrogen. Hum Reprod. 1998; 13:3000–4. https://doi.org/10.1093/humrep/13.11.3000 PMID: 9853845
2. Breedlove SM. Minireview: Organizational hypothesis: instances of the fingerpost. Endocrinology. 2010; 151(9):4116–22. https://doi.org/10.1210/en.2010-0041 PMID: 20631003
3. Valla J, Ceci SJ. Can Sex Differences in Science Be Tied to the Long Reach of Prenatal Hormones? Brain Organization Theory, Digit Ratio (2D/4D), and Sex Differences in Preferences and Cognition. Perspect Psychol Sci. 2011; 6(2):134–6. https://doi.org/10.1177/1745691611400236 PMID: 22164187
4. Wang Y, Wang HL, Li YH, Zhu FL, Li SJ, Ni H. Using 2D: 4D digit ratios to determine motor skills in children. European review for medical and pharmacological sciences. 2016; 20(5):806–9. PMID: 27010133
5. Pasanen BE, Tomkinson JM, Dufner TJ, Park CW, Fitzgerald JS, Tomkinson GR. The relationship between digit ratio (2D:4D) and muscular fitness: A systematic review and meta-analysis. Am J Hum Biol. 2021:e23657. https://doi.org/10.1002/ajhb.23657 PMID: 34331730
6. Siegmann EM, Müller T, Dziadeck I, Mühle C, Lenz B, Kornhuber J. Digit ratio (2D:4D) and transgender identity: new original data and a meta-analysis. Scientific reports. 2020; 10(1):19326. https://doi.org/10.1038/s41598-020-72486-6 PMID: 33168880
7. Cohen-Bendahan CC, van de Beek C, Berenbaum SA. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. Neurosci Biobehav Rev. 2005; 29(2):353–84. https://doi.org/10.1016/j.neubiorev.2004.11.004 PMID: 15811504
8. Hönekopp J, Watson S. Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand. Am J Hum Biol. 2010; 22(5):619–30. https://doi.org/10.1002/ajhb.21054 PMID: 20737609
9. Galis F, Ten Broek CM, Van Dongen S, Wijnaendts LC. Sexual dimorphism in the prenatal digit ratio (2D:4D). Arch Sex Behav. 2010; 39(1):57–62. https://doi.org/10.1007/s10508-009-9485-7 PMID: 19301112
10. Malas MA, Dogan S, Evcil EH, Desdicioglu K. Fetal development of the hand, digits and digit ratio (2D:4D). Early Hum Dev. 2006; 82(7):469–75. https://doi.org/10.1016/j.earlhumdev.2005.12.002 PMID: 16473482
11. Berenbaum SA, Bryk KK, Nowak N, Quigley CA, Moffat S. Fingers as a marker of prenatal androgen exposure. Endocrinology. 2009; 150(11):5119–24. https://doi.org/10.1210/en.2009-0774 PMID: 19819951
12. van Hemmen J, Cohen-Kettenis PT, Steensma TD, Veltman DJ, Bakker J. Do sex differences in CEOAEs and 2D:4D ratios reflect androgen exposure? A study in women with complete androgen insensitivity syndrome. Biol Sex Differ. 2017; 8:11. https://doi.org/10.1186/s13293-017-0132-z PMID: 28413602
13. Wallen K. Does Finger Fat Produce Sex Differences in Second to Fourth Digit Ratios? Endocrinology. 2009; 150(11):4819–22. https://doi.org/10.1210/en.2009-0986 PMID: 19846613
14. Brown WM, Hines M, Fane BA, Breedlove SM. Masculinized finger length patterns in human males and females with congenital adrenal hyperplasia. Horm Behav. 2002; 42(4):380–6. https://doi.org/10.1006/hbeh.2002.1830 PMID: 12488105

15. Ökten A, Kalyoncu M, Yariş N. The ratio of second- and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Early Hum Dev. 2002; 70(1–2):47–54. https://doi.org/10.1016/s0378-3782(02)00073-7 PMID: 12441204

16. Richards G, Browne WV, Aydin E, Constantinescu M, Nave G, Kim MS, et al. Digit ratio (2D:4D) and congenital adrenal hyperplasia (CAH): Systematic literature review and meta-analysis. Horm Behav. 2020; 126:104867. https://doi.org/10.1016/j.yhbeh.2020.104867 PMID: 32998030

17. Buck JJ, Williams RM, Hughes IA, Acerini CL. In-utero androgen exposure and 2nd to 4th digit length radio. Hum Reprod. 2003; 18(5):976–9.

18. Chang S, Skakkebæk A, Trolle C, Bojesen A, Hertz JM, Cohen A, et al. Anthropometry in Klinefelter Syndrome—Multifactorial Influences Due to CAG Length, Testosterone Treatment and Possibly Intrauterine Hypogonadism. The Journal of Clinical Endocrinology & Metabolism. 2015; 100(3):E508–E17. https://doi.org/10.1210/jc.2014-2834 PMID: 25514102

19. Manning JT, Kilduff LP, Trivers R. Digit ratio (2D:4D) in Klinefelter's syndrome. Andrology. 2013; 1 (1):94–9. https://doi.org/10.1111/j.2047-2927.2012.00013.x PMID: 23258636

20. Wong WI, Hines M. Interpreting digit ratio (2D:4D)-behavior correlations: 2D:4D sex difference, stability, and behavioral correlates and their replicability in young children. Horm Behav. 2016; 78:86–94. https://doi.org/10.1016/j.yhbeh.2015.10.022 PMID: 26542674

21. Hönekopp J. Digit ratio 2D:4D in relation to autism spectrum disorders, empathizing, and systemizing: a quantitative review. Autism research: official journal of the International Society for Autism Research. 2012; 5(4):221–30. https://doi.org/10.1002/aur.1230 PMID: 22674640

22. Fink B, Manning JT, Williams JHG, Podmore-Nappin C. The 2nd to 4th digit ratio and developmental psychopathology in school-aged children. Pers Individ Dif. 2007; 42(2):369–79.

23. Bunevicius A. The Association of Digit Ratio (2D : 4D) with Cancer: A Systematic Review and Meta-Analysis. Disease markers. 2018; 2018:7698193.

24. Fonseca CAD, Martelli DRB, Maia CMF, Dias VO, Carvalho AA, Júnior HM. Digital biomarker 2D:4D as a predictor of cancer: A systematic review. Early Hum Dev. 2022; 164:105521. https://doi.org/10.1016/j.earlhumdev.2021.105521 PMID: 34922146

25. Smoliga JM, Fogaca LK, Siplon JS, Goldburt AA, Jakobs F. Giving science the finger-is the second-to-fourth digit ratio (2D:4D) a biomarker of good luck? A cross sectional study. BMJ (Clinical research ed). 2021; 375:e067849. https://doi.org/10.1136/bmj-2021-067849 PMID: 34911738

26. Richards G. What is the evidence for a link between digit ratio (2D:4D) and direct measures of prenatal sex hormones? Early Hum Dev. 2017; 113:71–2. https://doi.org/10.1016/j.earlhumdev.2017.08.003 PMID: 28807670

27. Pruszkowska-Przybylska P, Sitek A, Rosset I, Sobalska-Kwapis M, Słomka M, Strapagiel D, et al. Association of the 2D:4D digit ratio with body composition among the Polish children aged 6–13 years. Early Hum Dev. 2018; 124:26–32.

28. Manning JT, Fink B, Mason L, Trivers R. Parental income inequality and children's digit ratio (2D:4D): a 'Trivers-Willard' effect on prenatal androgenization? Journal of biosocial science. 2022; 54(1):154–62. https://doi.org/10.1017/S0021932021000043 PMID: 33557976

29. Swift-Gallant A, Di Rita V, Major CA, Breedlove CJ, Jordan CL, Breedlove SM. Differences in digit ratios between gay men who prefer receptive versus insertive sex roles indicate a role for prenatal androgen. Scientific reports. 2021; 11(1):8102. https://doi.org/10.1038/s41598-021-87338-0 PMID: 33854100

30. Pruszkowska-Przybylska P, Sitek A, Rosset I, Sobalska-Kwapis M, Słomka M, Strapagiel D, et al. Associations between second to fourth digit ratio, cortisol, vitamin D, and body composition among Polish children. Scientific reports. 2021; 11(1):7029. https://doi.org/10.1038/s41598-021-86521-7 PMID: 33782473

31. Pruszkowska-Przybylska P, Kobus M, Iljin A, Wiktorska JA, Żądzińska E, Sitek A. Thyroid diseases and second to fourth digit ratio in Polish adults. Scientific reports. 2021; 11(1):18979. https://doi.org/10.1038/s41598-021-98436-4 PMID: 34556783

32. Kobus M, Sitek A, Antoszewski B, Rożniecki J, Pełka J, Żądzińska E. Prenatal oestrogen-testosterone balance as a risk factor of migraine in adults. The journal of headache and pain. 2021; 22(1):119. https://doi.org/10.1186/s10194-021-01326-3 PMID: 34620097

33. van de Beek C, Thijssen JH, Cohen-Kettenis PT, van Goozen SH, Buitelaar JK. 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? Horm Behav. 2004; 46 (5):663–9. https://doi.org/10.1016/j.yhbeh.2004.06.010 PMID: 15555509

34. Knickmeyer RC, Baron-Cohen S. Fetal testosterone and sex differences. Early Hum Dev. 2006; 82 (12):755–60. https://doi.org/10.1016/j.earlhumdev.2006.09.014 PMID: 17084045

35. Judd HL, Robinson JD, Young PE, Jones OW. Amniotic fluid testosterone levels in midpregnancy. Obstetrics and gynecology. 1976; 48(6):690–2. PMID: 999757

36. Robinson JD, Judd HL, Young PE, Jones OW, Yen SS. Amniotic fluid androgens and estrogens in midgestation. The Journal of clinical endocrinology and metabolism. 1977; 45(4):755–61. https://doi.org/10.1210/jcem-45-4-755 PMID: 144143

37. Abramovich DR. Human sexual differentiation—in utero influences. The Journal of obstetrics and gynaecology of the British Commonwealth. 1974; 81(6):448–53. https://doi.org/10.1111/j.1471-0528.1974.tb00494.x PMID: 4407079

38. Finegan JA, Bartleman B, Wong PY. A window for the study of prenatal sex hormone influences on postnatal development. The Journal of genetic psychology. 1989; 150(1):101–12. https://doi.org/10.1080/00221325.1989.9914580 PMID: 2496195

39. Di Mascio D, Khalil A, Rizzo G, Buca D, Liberati M, Martellucci CA, et al. Risk of fetal loss following amniocentesis or chorionic villus sampling in twin pregnancy: systematic review and meta-analysis. Ultrasound in Obstetrics & Gynecology. 2020; 56(5):647–55. https://doi.org/10.1002/uog.22143 PMID: 32632979

40. Akolekar R, Beta J, Picciarelli G, Ogilvie C, D'Antonio F. Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: a systematic review and meta-analysis. Ultrasound in Obstetrics & Gynecology. 2015; 45(1):16–26. https://doi.org/10.1002/uog.14636 PMID: 25042845

41. Nicolaides KH. First-trimester screening for chromosomal abnormalities. Seminars in perinatology. 2005; 29(4):190–4. https://doi.org/10.1053/j.semperi.2005.06.001 PMID: 16104667

42. Constantinescu M, Hines M. Relating Prenatal Testosterone Exposure to Postnatal Behavior in Typically Developing Children: Methods and Findings. Child Dev Perspect. 2012:n/a-n/a.

43. Lutchmaya S, Baron-Cohen S, Raggatt P, Knickmeyer R, Manning JT. 2nd to 4th digit ratios, fetal testosterone and estradiol. Early Hum Dev. 2004; 77(1–2):23–8. https://doi.org/10.1016/j.earlhumdev.2003.12.002 PMID: 15113628

44. Richards G, Browne W, Constantinescu M. Digit ratio (2D:4D) and amniotic testosterone and estradiol: An attempted replication of Lutchmaya et al. (2004). J Dev Orig Health Dis. 2020.

45. Manning JT. Digit Ratio: A Pointer to Fertility, Behavior, and Health. New Brunswick, New Jersey, London: Rutgers University Press; 2002.

46. Manning JT, Fink B, Mason L, Kasielska-Trojan A, Trivers R. The effects of sex, nation, ethnicity, age and self-reported pubertal development on participant-measured right-left 2D:4D (Dr-l) in the BBC internet study. Journal of biosocial science. 2022:1–13. https://doi.org/10.1017/S0021932022000049 PMID: 35088686

47. Manning J, Kilduff L, Cook C, Crewther B, Fink B. Digit Ratio (2D:4D): A Biomarker for Prenatal Sex Steroids and Adult Sex Steroids in Challenge Situations. Frontiers in endocrinology. 2014; 5. https://doi.org/10.3389/fendo.2014.00009 PMID: 24523714

48. Ventura T, Gomes MC, Pita A, Neto MT, Taylor A. Digit ratio (2D:4D) in newborns: influences of prenatal testosterone and maternal environment. Early Hum Dev. 2013; 89(2):107–12. https://doi.org/10.1016/j.earlhumdev.2012.08.009 PMID: 23017880

49. Ernsten L, Körner LM, Heil M, Richards G, Schaal NK. Investigating the reliability and sex differences of digit lengths, ratios, and hand measures in infants. Scientific reports. 2021; 11(1):10998. https://doi.org/10.1038/s41598-021-89590-w PMID: 34040007

50. Trivers R, Manning J, Jacobson A. A longitudinal study of digit ratio (2D:4D) and other finger ratios in Jamaican children. Horm Behav. 2006; 49:150–6. https://doi.org/10.1016/j.yhbeh.2005.05.023 PMID: 16040033

51. McIntyre MH, Ellison PT, Lieberman DE, Demerath E, Towne B. The development of sex differences in digital formula from infancy in the Fels Longitudinal Study. Proc R Soc Lond B Biol Sci. 2005; 272 (1571):1473–9. https://doi.org/10.1098/rspb.2005.3100 PMID: 16011922

52. McIntyre MH, Cohn BA, Ellison PT. Sex dimorphism in digital formulae of children. Am J Phys Anthropol. 2006; 129(1):143–50. https://doi.org/10.1002/ajpa.20240 PMID: 16224778

53. Knickmeyer RC, Woolson S, Hamer RM, Konneker T, Gilmore JH. 2D:4D ratios in the first 2 years of life: Stability and relation to testosterone exposure and sensitivity. Horm Behav. 2011; 60(3):256–63. https://doi.org/10.1016/j.yhbeh.2011.05.009 PMID: 21664359

54. Williams JHG, Greenhalgh KD, Manning JT. Second to fourth finger ratio and possible precursors of developmental psychopathology in preschool children. Early Hum Dev. 2003; 72(1):57–65. https://doi.org/10.1016/s0378-3782(03)00012-4 PMID: 12706312

55. Richards G, Bellin W, Davies W. Familial digit ratio (2D:4D) associations in a general population sample from Wales. Early Hum Dev. 2017; 112:14–9. https://doi.org/10.1016/j.earlhumdev.2017.06.006 PMID: 28668648

56. Manning JT, Fink B. Sexual dimorphism in the ontogeny of second (2D) and fourth (4D) digit lengths, and digit ratio (2D:4D). Am J Hum Biol. 2018; 30(4):e23138. https://doi.org/10.1002/ajhb.23138 PMID: 30133905

57. Matzdorff I. Metrische und allometrische Wachstumsuntersuchungen an der menschlichen Hand. Z Morphol Anthropol. 1967; 59(2):158–84.

58. Körner LM, Pause BM, Meinlschmidt G, Tegethoff M, Frohlich S, Kozlowski P, et al. Prenatal testosterone exposure is associated with delay of gratification and attention problems/overactive behavior in 3-year-old boys. Psychoneuroendocrinology. 2019; 104:49–54. https://doi.org/10.1016/j.psyneuen.2019.02.014 PMID: 30802710

59. Körner LM, Schaper ML, Pause BM, Heil M. Parent-Reports of Sex-Typed Play Preference in Preschool Children: Relationships to 2D:4D Digit Ratio and Older Siblings' Sex. Arch Sex Behav. 2020; 49(7):2715–24. https://doi.org/10.1007/s10508-020-01662-6 PMID: 32222854

60. Kemper CJ, Schwerdtfeger A. Comparing indirect methods of digit ratio (2D:4D) measurement. Am J Hum Biol. 2009; 21(2):188–91. https://doi.org/10.1002/ajhb.20843 PMID: 18988284

61. Mikac U, Buško V, Sommer W, Hildebrandt A. Analysis of different sources of measurement error in determining second-to-fourth digit ratio, a potential indicator of perinatal sex hormones exposure. Annu Rev Psychol. 2016; 23:39–49.

62. DeBruine LM. AutoMetric software for measurement of 2D:4D ratios. www.facelab.org/debruine/Programs/autometric.php2004.

63. Krull J, MacKinnon D. Multilevel Modeling of Individual and Group Level Mediated Effects. Multivariate Behav Res. 2001; 36:249–77. https://doi.org/10.1207/S15327906MBR3602_06 PMID: 26822111

64. Kenny DA, Korchmaros JD, Bolger N. Lower level mediation in multilevel models. Psychol Methods. 2003; 8(2):115–28. https://doi.org/10.1037/1082-989x.8.2.115 PMID: 12924810

65. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015; 67(1):1–48.

66. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest: Tests for random and fixed effects for linear mixed effect models [Software]. http://cran.r-project.org/package=lmerTest 2014.

67. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

68. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2 ed: Lawrence Erlbaum Associates; 1988.

69. Kuijper EAM, Ket JCF, Caanen MR, Lambalk CB. Reproductive hormone concentrations in pregnancy and neonates: a systematic review. Reprod Biomed Online. 2013; 27(1):33–63. https://doi.org/10.1016/j.rbmo.2013.03.009 PMID: 23669015

70. Hollier LP, Keelan JA, Hickey M, Maybery MT, Whitehouse AJO. Measurement of Androgen and Estrogen Concentrations in Cord Blood: Accuracy, Biological Interpretation, and Applications to Understanding Human Behavioral Development. Frontiers in endocrinology. 2014; 5(64). https://doi.org/10.3389/fendo.2014.00064 PMID: 24829559

71. Hickey M, Doherty DA, Hart R, Norman RJ, Mattes E, Atkinson HC, et al. Maternal and umbilical cord androgen concentrations do not predict digit ratio (2D:4D) in girls: A prospective cohort study. Psychoneuroendocrinology. 2010; 35(8):1235–44. https://doi.org/10.1016/j.psyneuen.2010.02.013 PMID: 20299156

72. Barrett E, Thurston SW, Harrington D, Bush NR, Sathyanarayana S, Nguyen R, et al. Digit ratio, a proposed marker of the prenatal hormone environment, is not associated with prenatal sex steroids, anogenital distance, or gender-typed play behavior in preschool age children. J Dev Orig Health Dis. 2020:1–10.

73. Richards G, Gomes M, Ventura T. Testosterone measured from amniotic fluid and maternal plasma shows no significant association with directional asymmetry in newborn digit ratio (2D:4D). J Dev Orig Health Dis. 2019; 10(3):362–7. https://doi.org/10.1017/S2040174418000752 PMID: 30376903

74. Richards G, Aydin E, Tsompanidis A, Padaigaitė E, Austin T, Allison C, et al. Digit ratio (2D:4D) and maternal testosterone-to-estradiol ratio measured in early pregnancy. Scientific reports. 2022; 12(1):13586. https://doi.org/10.1038/s41598-022-17247-3 PMID: 35945232

75. Swift-Gallant A, Johnson BA, Di Rita V, Breedlove SM. Through a glass, darkly: Human digit ratios reflect prenatal androgens, imperfectly. Horm Behav. 2020; 120:104686. https://doi.org/10.1016/j.yhbeh.2020.104686 PMID: 32014464

76. Zheng Z, Cohn MJ. Developmental basis of sexually dimorphic digit ratios. Proc Natl Acad Sci U S A. 2011; 108(39):16289–94. https://doi.org/10.1073/pnas.1108312108 PMID: 21896736

77. Huber SE, Lenz B, Kornhuber J, Müller CP. Prenatal androgen-receptor activity has organizational morphological effects in mice. PloS one. 2017; 12(11):e0188752. https://doi.org/10.1371/journal.pone.0188752 PMID: 29176856

78. Swift-Gallant A, Di Rita V, Coome LA, Monks DA. Overexpression of Androgen Receptors Masculinizes 2D:4D Digit Ratios in Mice. Arch Sex Behav. 2021. https://doi.org/10.1007/s10508-021-02161-y PMID: 34625883

79. Talarovičová A, Kršková L, Blažeková J. Testosterone enhancement during pregnancy influences the 2D:4D ratio and open field motor activity of rat siblings in adulthood. Horm Behav. 2009; 55(1):235–9. https://doi.org/10.1016/j.yhbeh.2008.10.010 PMID: 19022257

80. Suchonova M, Borbelyova V, Renczes E, Konecna B, Vlkova B, Hodosy J, et al. Does the 2nd and 4th digit ratio reflect prenatal androgen exposure? Bratislavske lekarske listy. 2019; 120(9):703–10. https://doi.org/10.4149/BLL_2019_119 PMID: 31475559

81. McIntyre MH. The use of digit ratios as markers for perinatal androgen action. Reproductive biology and endocrinology: RB&E. 2006; 4:10. https://doi.org/10.1186/1477-7827-4-10 PMID: 16504142

82. Manning JT. Sex Differences and Age Changes in Digit Ratios: Implications for the Use of Digit Ratios in Medicine and Biology. In: Preedy VR, editor. Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease. New York, NY: Springer New York; 2012. p. 841–51.

83. Berenbaum SA, Beltz AM. Sexual differentiation of human behavior: effects of prenatal and pubertal organizational hormones. Front Neuroendocrinol. 2011; 32(2):183–200. https://doi.org/10.1016/j.yfrne.2011.03.001 PMID: 21397624

84. Auyeung B, Lombardo MV, Baron-Cohen S. Prenatal and postnatal hormone effects on the human brain and cognition. Pflugers Arch. 2013; 465(5):557–71. https://doi.org/10.1007/s00424-013-1268-2 PMID: 23588379

85. Lamminmaki A, Hines M, Kuiri-Hanninen T, Kilpelainen L, Dunkel L, Sankilampi U. Testosterone measured in infancy predicts subsequent sex-typed behavior in boys and in girls. Horm Behav. 2012; 61(4):611–6. https://doi.org/10.1016/j.yhbeh.2012.02.013 PMID: 22373494

86. Manning JT. Digit ratio (2D:4D), sex differences, allometry, and finger length of 12-30-year olds: evidence from the British Broadcasting Corporation (BBC) Internet study. Am J Hum Biol. 2010; 22(5):604–8. https://doi.org/10.1002/ajhb.21051 PMID: 20737606

87. Kuzawa CW. Adipose tissue in human infancy and childhood: An evolutionary perspective. Am J Phys Anthropol. 1998; 107(S27):177–209. https://doi.org/10.1002/(sici)1096-8644(1998)107:27+<177::aid-ajpa7>3.0.co;2-b PMID: 9881526

88. Ribeiro E, Neave N, Morais RN, Manning JT. Direct Versus Indirect Measurement of Digit Ratio (2D:4D): A Critical Review of the Literature and New Data. Evolutionary psychology: an international journal of evolutionary approaches to psychology and behavior. 2016; 14(1):1474704916632536.

89. Burton L, Carachi R. The Limbs. In: Carachi R, Helmi S, Doss E, editors. Clinical Embryology: Springer Cham; 2019. p. 449–61.

90. Doss SH, Sneddon SF. General Embryology. In: Carachi R, Helmi S, Doss E, editors. Clinical Embryology: Springer Cham; 2019. p. 11–30.

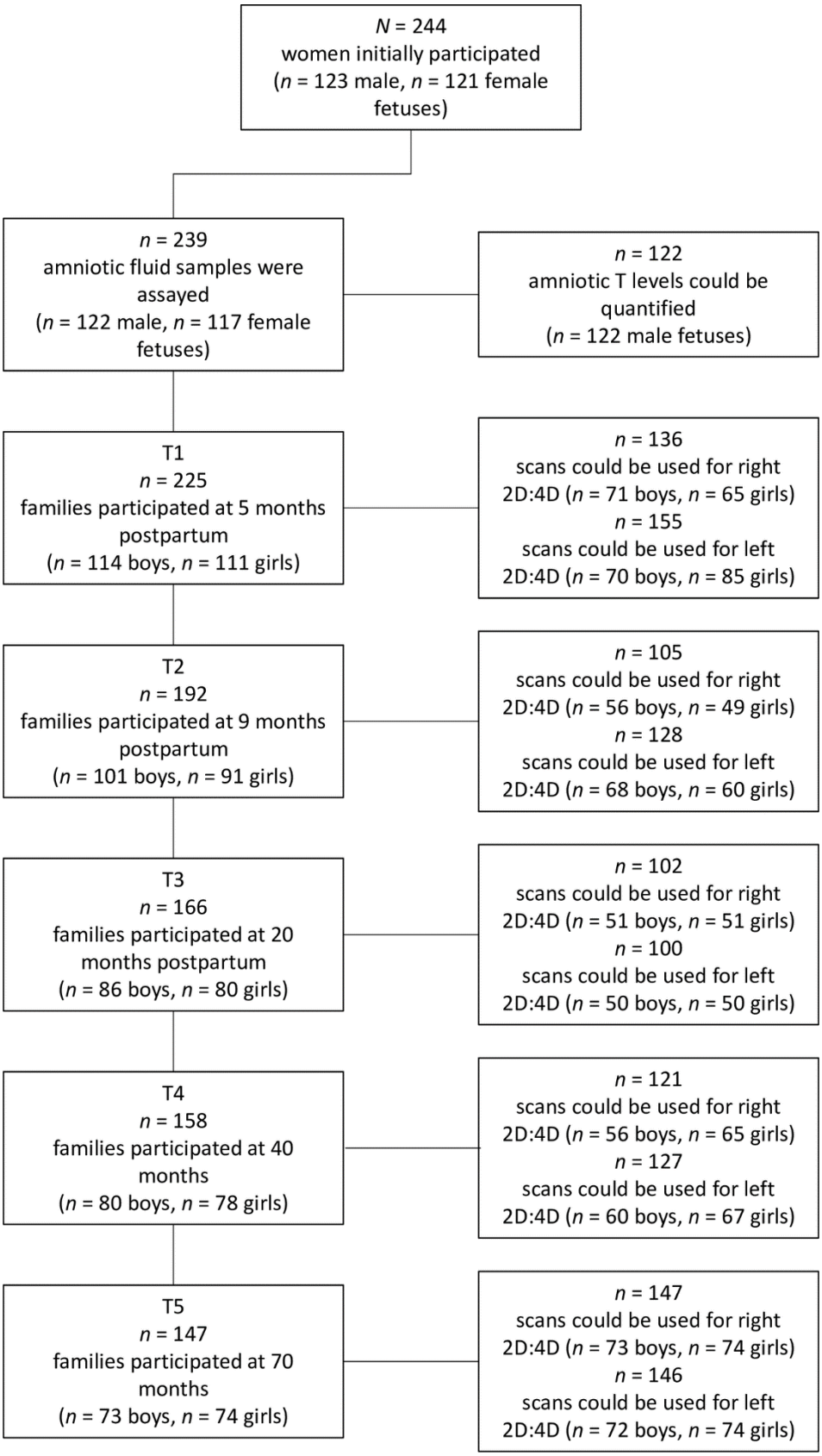
91. Navaratnam K, Alfirevic Z, Obstetricians tRCo, Gynaecologists. Amniocentesis and chorionic villus sampling. BJOG: An International Journal of Obstetrics & Gynaecology. 2022; 129(1):e1–e15.

92. Králík M, Hupková A, Zeman T, Hložek M, Hlaváček L, Slováčková M, et al. Family effects on the digit ratio (2D:4D): The role of the interbirth interval. American Journal of Human Biology. 2019; 31(4):e23260. https://doi.org/10.1002/ajhb.23260 PMID: 31183942

93. Saino N, Leoni B, Romano M. Human digit ratios depend on birth order and sex of older siblings and predict maternal fecundity. Behavioral Ecology and Sociobiology. 2006; 60(1):34–45.

94. Manning JT, Churchill AJG, Peters M. The Effects of Sex, Ethnicity, and Sexual Orientation on Self-Measured Digit Ratio (2D:4D). Archives of sexual behavior. 2007; 36(2):223–33. https://doi.org/10.1007/s10508-007-9171-6 PMID: 17373585

95. Butovskaya M, Burkova V, Apalkova Y, Dronova D, Rostovtseva V, Karelin D, et al. Sex, population origin, age and average digit length as predictors of digit ratio in three large world populations. Scientific reports. 2021; 11(1):8157. https://doi.org/10.1038/s41598-021-87394-6 PMID: 33854119



S Figure 1

Data inclusion and exclusion chart.

Study 2:

Ernsten, L., Körner, L. M., Heil, M., Richards, G., & Schaal, N. K. (2021). Investigating the reliability and sex differences of digit lengths, ratios, and hand measures in infants. *Sci Rep*, *11*(1), 10998. https://doi.org/10.1038/s41598-021-89590-w

Author contributions: Ernsten, L.: Conceptualization; Data curation, Formal analysis, Methodology, Project administration, Visualization, Writing – original draft; Körner, L. M.: Conceptualization, Methodology, Writing – review & editing; Heil, M.: Conceptualization, Methodology, Investigation, Supervision, Writing – Review & Editing; Richards, G.: Conceptualization, Methodology, Writing – Review & Editing; Schaal, N. K.: Conceptualization, Methodology, Supervision, Writing – Review & Editing

Impact Factor (2022): 4.6 (Journal Citation Reports, 2023c)

scientific reports

Check for updates

OPEN

Investigating the reliability and sex differences of digit lengths, ratios, and hand measures in infants

Luisa Ernsten[1]✉, Lisa M. Körner[1], Martin Heil[1], Gareth Richards[2] & Nora K. Schaal[1]

Hands and digits tend to be sexually dimorphic and may reflect prenatal androgen exposure. In the past years, the literature introduced several hand and digit measures, but there is a lack of studies in prepubertal cohorts. The available literature reports more heterogeneous findings in prepubertal compared to postpubertal cohorts. The comparability of the available studies is further limited by the study design and different measurement techniques. The present study compared the reliability and sex differences of available hand and digit measures, namely digit lengths of 2D, 3D, 4D, 5D, digit ratios 2D:4D, 2D:5D, 3D:4D, 3D:5D, 4D:5D, relative digit lengths rel2, rel3, rel4, rel5, directional asymmetry of right and left 2D:4D (D_{r-l}), hand width, length, and index of 399 male and 364 female 6-month-old German infants within one study using only indirect and computer-assisted measurements. The inter-examiner reliability was excellent while the test-retest reliability of hand scans was only moderate to high. Boys exhibited longer digits as well as wider and longer hands than girls, but smaller digit ratios, with ratios comprising the fifth digit revealing the largest effect sizes. Other hand and digit ratios revealed sex differences to some extent. The findings promote the assumption of sexual dimorphic hand and digit measures. However, by comparing the results of the available literature, there remains an uncertainty regarding the underlying hypothesis. Specifically in prepubertal cohorts, i.e. before the influence of fluctuating hormones, significant effects should be expected. It seems like other factors than the influence of prenatal androgens contribute to the sexual dimorphism in hand and digit lengths.

Sexual determination and differentiation can be observed in males and females of sexually reproducing species. Its origins lie in complex relations of biology, genetics, and social as well as physical environments[1,2]. While genetics primarily determine the gonads in the offspring, sex hormones like androgens and estrogens secondarily promote the phenotypic differentiation of males and females[3]. One of those phenotypic differences between male and female humans is the hand, as males exhibit generally bigger hands compared to females[4,5].

In the past years, research suggested that hands and digits may serve as an indicator for sexual differentiation, as it is hypothesized that they are associated to the HOX genes[6], which promote the development of the urogenital tract and external genitalia as well as the limb development[7,8], and reflect differences in prenatal androgen exposure[9]. Therefore, there is extensive research on different hand and digit measures that are hypothesized to reflect different influences of prenatal androgens[10]. However, most studies regarding hand and digit measures are carried out with adult samples[11,12], albeit the investigation of those ratios in young cohorts before onset of puberty, i.e. the effect of fluctuating hormones, seems intriguing as one can assume that sex and gender differences in this period are mainly associated to organizational effects of prenatal sex determining and differentiating factors such as genetics and sex hormones[13–15]. The literature and reported effects in hand and digit measures used as a marker for prenatal sex hormone exposure seem to be more homogeneous in adult samples[11], while in prepubertal children there is considerable heterogeneity and age-dependent fluctuation regarding robust effects[16,17]. Yet, findings regarding different digit or hand measures as markers of prenatal androgens in younger cohorts are sparse.

Different hand and digit measures were discussed in the past research and shall be briefly presented (for an overview see supplementary table 1). The most discussed indicator for prenatal androgen exposure is the second to fourth digit ratio (2D:4D) that has also been associated with an amount of behavioral outcomes that are known to differ between males and females[18]. By comparing the length of the second digit to the length

[1]Department of Experimental Psychology, Heinrich-Heine-University, Universitätsstraße 1, 40225 Düsseldorf, Germany. [2]School of Psychology, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. ✉email: Luisa.Ernsten@hhu.de

of the fourth digit, the general assumption of a lower 2D:4D in males compared to females has been revealed in many studies[11,17,19]. A sex difference in 2D:4D can be found as early as 14 weeks of gestational age[19] but its stability over time varies across different studies[16,20]. Besides 2D:4D, the literature aimed to evaluate other hand and digit measures in humans as possible alternative markers for prenatal androgen exposure[21]. By computing digit ratios with every possible digit (excluding the thumb), i.e. the second (2D), third (3D), fourth (4D), and fifth digit (5D), significant differences between male and female ratios in different age groups could be observed in various studies[22–26]. However, a study regarding directly measured digit ratios in children aged 2–18 years revealed relevant fluctuations and argues that digit ratios other than 2D:4D do not serve as reliable indicators of prenatal androgen exposure[12]. On the other hand, the studies of Dressler & Voracek (2011) and Kumar et al. (2017) reported only small[26] or no effects in 2D:4D[22] and suggested that digit ratios with 5D as one of the components may reveal larger sex differences compared to 2D:4D[22]. But it is important to note that these studies used different measurement techniques. Whereas Kumar and colleagues (2017) measured the digit length dorsally[22], most studies define the digit length over landmarks and flexion creases derived from the ventral surface of the hand[12,27]. Therefore, another study that used radiographs of children aged one month to 18 years and gives evidence for robust sex differences in different digit ratios[28] might be difficult to compare to the aforementioned studies using digit lengths derived from measuring the soft tissue[12,22–26]. Another introduced measure is the relative digit length: By dividing the length of one digit by that of the sum of all four digits taken together, the contribution of each digit concerning a computed digit ratio can be determined[29]. It has been shown that there are sex differences in those relative digit lengths and that males in general exhibit larger relative lengths of 4D and 5D, while females exhibit larger relative lengths of 2D and 3D with low to medium effect sizes[29,30]. To the best of our knowledge, no study has considered the relative digit lengths in prepubertal cohorts. Another aspect to consider when looking at sexual dimorphism of the hand is that authors postulate that prenatal androgens have a different influence on the right versus left side of the body and digit ratios may differ in the right versus left hand[11,31,32]. In order to examine this assumption, it was Manning (2002) who introduced a measure reflecting directional asymmetry by subtracting left 2D:4D from right 2D:4D (D_{r-l})[32]. Values < 0 indicate a right-biased asymmetry (i.e. 2D:4D is lower in the right relative to the left hand) which reflects high levels of early androgen exposure[33]. However, relatively little research has considered this variable in terms of sexual dimorphism and D_{r-l} in newborns remained uncorrelated towards testosterone levels in amniotic fluid and showed no significant sex difference[34].

The hand itself as a discriminator between the sexes is widely used in forensic contexts to determine the sex of dismembered human extremities[35]. It could be shown that males tend to have larger and wider hands compared to females and also the hand index differs between the sexes[35–37]. Sex differences in hands are relatively stable[5], however, in the fetal age no significant sex differences in hand length, width, and index [i.e. (Hand width/Hand length)*100] could be observed in radiographs of 50 fetuses between 20 and 40 weeks of gestational age[38] and also in prepubertal children a considerable overlap between males and females considering the size of the hand remains[39]. With onset of the puberty, sex differences can be reliably observed[5]. A differentiation between males and females on the basis of hand measures may be more difficult in younger, i.e. prepubertal, cohorts compared to adolescent or adult samples. Albeit one can assume that the development of the hands relies on the same mechanisms as of digits, to the best of our knowledge, no study has investigated hand width, length, or index in the context of prenatal androgen exposure.

Despite considerable evidence for sexual dimorphic hand and digit measures, it is to note that studies differ in terms of their study design and measurement method. There are differences concerning the dimension of the hand from which lengths are examined (ventral[20,26,29] versus dorsal[22,36]), the measurement technique (direct[17,24,30,36,37] versus indirect[16,17,20,25,29]) and the used tools for measurement (computer based[25,29] versus caliper[16,17,22,24,26,30,36,37] versus ruler[17,19]) as well as the person examining the lengths (examiner[16,17,19,20,22,30] versus self-measured[17,29]), or even the source for computing digit lengths, i.e. using radiographs[19,21] rather than soft tissue[16,17,20,22,24–26,29,30,37](see supplementary table 1). Comparing these different measurement techniques, they differ in terms of precision and can produce considerable variability in examinations and statistical effects[35,40,41].

In sum, the literature provides valuable information about sex differences in hand and digit measures and it is hypothesized that the sexual dimorphic growth pattern is associated to the influence of genetics and prenatal androgen action. However, there are two main concerns regarding the available literature: (1) it appears that there is a general lack of studies examining and comparing different hand and digit ratios in younger, specifically in prepubertal cohorts. This is especially relevant as hand and digit growth appears to underlie age-dependent variations and a robust sex difference emerges with onset of puberty. And (2) the comparability of different studies is limited due to different measurement techniques used to assess hand and digit measures. The current study aims to compare hand and digit measures which have been already introduced in the literature, namely digit lengths of 2D, 3D, 4D, 5D, digit ratios of 2D:4D, 2D:5D, 3D:4D, 3D:5D, 4D:5D, relative digit lengths of 2D, 3D, 4D, 5D, as well as hand width, length, and index in a sample of infants ($N = 763$) within one study using the same measurement technique for each measure. We solely rely on indirect measurements using hand-scans and a computer-program as it is proposed that this technique shows the highest precision[40,42]. Based on previous studies, a sex difference between digit lengths, 2D:4D and other digit ratios is expected, and it is assumed that boys exhibit larger digit lengths but smaller digit ratios than girls. Furthermore, relative digit lengths, directional asymmetry in right and left 2D:4D (D_{r-l}), and the hand width, length, and index are also investigated. Furthermore, we aim to investigate the reliability of repeated measurements and inter-examiner reliability to give valuable information on methodological considerations in examining hand and digit measures in very young cohorts.

Methods

Participants. Families with newborn children were recruited as part of other studies conducted between 2013 and 2018 at the Department of Experimental Psychology at Heinrich-Heine-University Düsseldorf. All parents spoke German fluently and almost all infants were White and from middle-class backgrounds. They were invited to take part in infant studies when their child was 6 months of age. In total 1381 (702 boys and 679 girls) infants with a mean age of 195.18 days ($SD = 8.40$) participated.

Procedure. Families with 6-month-old infants came to the Department of Experimental Psychology at Heinrich-Heine-University Düsseldorf in order to take part in a mental rotation experiment [see[43]]. Informed consent was obtained from all parents and/or legal guardians of participating infants. After performing the mental rotation task, examiners took hand scans from a total of 1381 infants by pushing the ventral surface of the infant's hand lightly onto the scanner glass and covering it up with a towel. Due to unexpected movements of the infants, particular digits could not be measured correctly. For data analysis, only hand scans with measurable 2D, 3D, 4D, and 5D were used ($N = 790$). In addition, outliers > 2 interquartile range (IQR) from the median were excluded ($N = 27$) resulting in a final sample of 763 infants/hand scans. The sample consisted of 399 boys and 364 girls. Of this sample, 180 scans (56% male) could be used for analysis of the hand length, width and index. A majority of the scans provided no valid measurement of hand length and width because the landmarks for those measures were covered by clothing or the thumb. Furthermore, for a subsample of 130 children (61% male), that were obtained as the last cohorts, scans were taken twice, i.e. at the beginning as well as at the end of the study to test for reliability of the hand scan measures. The families received a refund of their travel expenses, and the study was approved by the ethics committee of the Heinrich-Heine-University Düsseldorf in Germany.

Measures. Scans of both right and left hand were obtained by an examiner using a FUJITSU fi-60F image scanner and digit lengths were measured using the freeware program *AutoMetric*[44]. Digital scans can be uploaded in the program and for each digit, excluding the thumb, the length can be determined by first clicking at the tip of each digit and second clicking at the midpoint of the ventral proximal flexion crease of each digit (see Fig. 1). The program then automatically computes the digit length into pixels. A monitor with 100 dpi was used, which means 100 pixels relate to 2.54 cm. The program was also used for measuring the hand width and length (see Table 1 and Fig. 1). All digit measures, i.e. ratios, relative digit lengths, the difference between right and left 2D:4D, and hand index, were derived from lengths measured in pixels and calculated using the statistical software *SPSS*[45] (see Table 1). As shown for 2D:4D by Ribeiro, Neave, Morais and Manning (2016)[46], indirect measures like hand scans produce larger sex differences and a higher measurement precision than direct measures; moreover, *AutoMetric* shows a high reliability for digit measurements and is superior to other computer-based measurement techniques[27,40]. Two independent examiners measured the hand scans and were blind to the sex of the children. Table 1 lists all hand and digit dimensions used and their computation base.

Statistical analysis. For the reliability analysis of the measurements, the inter-examiner reliability is estimated by intra-class correlations (ICC) between the two independent examiners of the hand scans. The ICC is interpreted as follows: $ICC < 0.50$ means poor, $0.50 < ICC < 0.75$ means fair, $0.75 < ICC < 0.90$ means good, and $ICC > 0.90$ means excellent inter-examiner reliability[47]. Furthermore, for the subsample with additional hand scans taken before the experiment, the reliability of the measurements of hand scans taken before and at the end of the experiment (hereinafter referred to as test-retest-reliability) is estimated via Pearson correlations for every measure on both hands and averaged over both hands separated for boys, girls, and averaged over both sexes. Pearson correlations were interpreted as follows: low correlation $r \geq 0.10$, moderate correlation $r \geq 0.30$, and high correlation $r \geq 0.50$[48].

To test for main effects and interactions between the factors *sex* (male vs. female), *hand* (right vs. left), and *digit measure* (see Table 1 that lists different digit and hand measures/ factors and the factor levels), three $2 \times 2 \times n$ (n for different factor levels) mixed ANOVAs with the between-subjects factor *sex*, and the within-subject factors *hand* and *digit measure* were conducted for digit length, digit ratio and relative digit length. In order to further disentangle the effects for sex, $2 \times n$ mixed ANOVAs were computed separately for each hand with the between-subjects factor *sex* and the within-subject factor *digit measure*. For the hand width, length, and index, three 2×2 mixed ANOVAs with the within-subject factor *hand* and the between-subjects factor *sex* were conducted. To further investigate which digit measure was influenced by sex, post-hoc independent samples *t*-tests between boys and girls for each digit, digit ratio and hand measure were computed. The focus of the analyses is on sex differences. We report the results of the ANOVA for the sake of completeness, however, only the main effects of the factor sex are further investigated. The sex difference in right-left asymmetry in 2D:4D ($D_{r\text{-}l}$) was tested using independent samples *t*-test. Multiple corrections were not applied, however, we report effects sizes *d* and their confident intervals as we believe that these measures promote the validity of our results.

Alpha-levels were set to 0.05 for each analysis, $\alpha \leq 0.10$ was interpreted as a statistical trend. Effect sizes are reported as η^2_p and interpreted as $\eta^2_p \geq 0.01$ small effect, $\eta^2_p \geq 0.06$ medium effect, and $\eta^2_p \geq 0.14$ large effect[48], or were converted to Cohen's *d* and interpreted according to Cohen[48]— small effect $d \geq 0.20$, medium effect $d \geq 0.50$, and large effect $d \geq 0.80$. *Greenhouse Geisser* adjustment was used to correct for violations of sphericity. All analyses were performed with *SPSS* version 27.0.

Ethics declaration. The research was conducted in accordance with the Declaration of Helsinki.

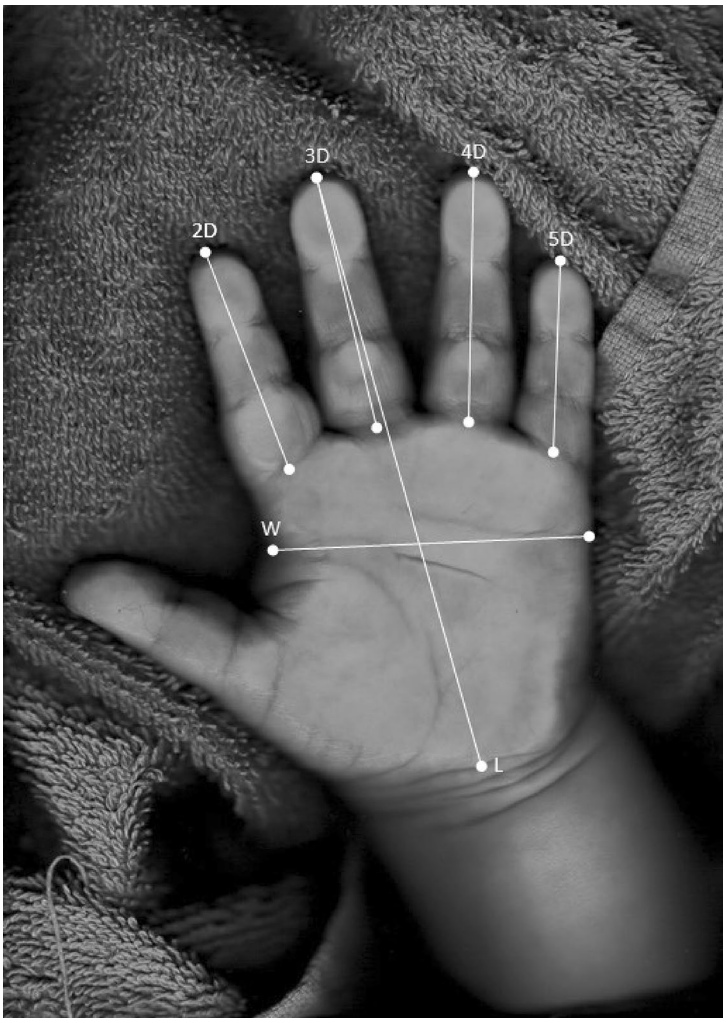


Figure 1. Example measurement of digit lengths (2D, 3D, 4D, 5D) and hand width (W) and length (L) using the freeware program *AutoMetric*[44].

Results

Reliability analysis. Intra-class correlations for the different digit lengths varied between 0.94 and 0.97 and were highly significant (all $p < 0.001$). Intra-class correlations for the hand width and length varied between 0.86 and 0.96 and also were highly significant (all $p < 0.001$). In supplementary table 2 and supplementary table 3 the computed intra-class correlations as well as their confidence intervals can be seen. The two measurements were averaged for each digit and hand measure on both hands to increase reliability. The following analyses were conducted with the averaged ratings.

The test-retest reliabilities of digit lengths are presented in supplementary table 4. They were significant for each digit and can be considered moderate to high. The reliabilities for the calculated digit measures were lower compared to those for digit lengths but can be also considered moderate to high. Scans taken at the beginning of the experiment that should be used for analyzing the hand width, length, and index were only available for 19 children (for the same reason as described above under *Procedure*) so that a reliability analysis could not be performed.

Digit lengths. A 2 × 2 × 4 mixed ANOVA revealed significant main effects of the factor *hand*, F (1, 761) = 21.67, $p < 0.001$, $\eta^2_p = 0.03$, the factor *digit measure*, F (2.30, 1,746.94) = 16,222.59, $p < 0.001$, $\eta^2_p = 0.96$, and the factor *sex*, F (1, 761) = 86.28, $p < 0.001$, $\eta^2_p = 0.10$, with digit lengths as the dependent variable. Furthermore, the *digit measure*sex* interaction, F (2.30, 1,746.94) = 6.41, $p = 0.001$, $\eta^2_p = 0.01$, and the *hand*digit measure* interaction, F (2.28, 1,734.93) = 30.84, $p < 0.001$, $\eta^2_p = 0.04$, were significant. The *hand*sex* and *hand*digit measure*sex* interactions were both non-significant ($p \geq 0.171$). A 2 × 4 mixed ANOVA separately conducted for the right and left hand showed significant main effects of the factor *digit measure*, F (2.25, 1,710.06) = 11,892.09, $p < 0.001$, $\eta^2_p = 0.94$, and the factor *sex*, F (1761) = 87.42, $p < 0.001$, $\eta^2_p = 0.10$, on digit lengths and a significant *digit measure*sex* interaction, F (2.25, 1710.06) = 4.20, $p = 0.012$, $\eta^2_p = 0.01$, for the right hand. For the left hand, there were significant main effects of the factor *digit measure*, F (2.37, 1803.36) = 11,991.80, $p < 0.001$, $\eta^2_p = 0.94$,

Digit length		
2D	Length of the index finger.	
3D	Length of the long finger.	
4D	Length of the annular finger.	
5D	Length of the auricular finger.	
Digit ratios		
2D:4D	Ratio of index finger to annular finger:	$2D:4D = \frac{2D}{4D}$
2D:5D	Ratio of index finger to auricular finger:	$2D:5D = \frac{2D}{5D}$
3D:4D	Ratio of long finger to annular finger:	$3D:4D = \frac{3D}{4D}$
3D:5D	Ratio of long finger to auricular finger:	$3D:5D = \frac{3D}{5D}$
4D:5D	Ratio of annular finger to auricular finger:	$4D:5D = \frac{4D}{5D}$
Relative digit lengths		
rel2	Ratio of index finger to all four fingers:	$rel2 = \frac{2D}{(2D+3D+4D+5D)}$
rel3	Ratio of long finger to all four fingers:	$rel3 = \frac{3D}{(2D+3D+4D+5D)}$
rel4	Ratio of annual finger to all four fingers:	$rel4 = \frac{4D}{(2D+3D+4D+5D)}$
rel5	Ratio of auricular finger to all four fingers:	$rel5 = \frac{5D}{(2D+3D+4D+5D)}$
Right-Left asymmetry in 2D:4D		
$D_{r\text{-}l}$	Difference between right and left 2D:4D:	$D_{r-l} = right\ 2D:4D - left\ 2D:4D$
Hand measures		
Width	Length measured by the ventral crease above the head of the second metacarpal bone to the ventral crease of the head of the fifth metacarpal bone.	
Length	Length measured by the tip of the long finger to the midpoint of the ventral distal crease of the carpus.	
Index	Ratio of hand width to hand length multiplied by 100:	$Index = \frac{Hand\ width}{Hand\ length} \times 100$

Table 1. Overview of calculated digit and hand measures.

and the factor *sex*, $F(1763) = 73.28$, $p < 0.001$, $\eta^2_p = 0.62$, on digit lengths and a significant *digit measure*sex* interaction, $F(2.37, 1803.36) = 5.83$, $p = 0.002$, $\eta^2_p = 0.01$. Post-hoc *t*-tests revealed significant differences between girls and boys in every digit length (all $p < 0.001$, see Table 2) with girls having shorter digits than boys. The means and standard deviations are presented in Fig. 2.

Digit ratios. The 2 × 2 x 5 mixed ANOVA showed significant main effects of the factors *hand*, $F(1, 761) = 44.77$, $p < 0.001$, $\eta^2_p = 0.06$, *digit measure*, $F(1.95, 1481.67) = 16{,}493.65$, $p < 0.001$, $\eta^2_p = 0.96$, and *sex*, $F(1761) = 31.32$, $p < 0.001$, $\eta^2_p = 0.04$, on digit ratios. Interactions were significant for *digit measure*sex*, $F(1.95, 1481.67) = 14.39$, $p < 0.001$, $\eta^2_p = 0.02$, and for *hand*digit measure*, $F(2.28, 1733.48) = 15.04$, $p < 0.001$, $\eta^2_p = 0.02$. The *hand*sex* and *hand*digit measure*sex* interactions were not significant (both $p \geq 0.166$). Separate 2 × 5 mixed ANOVA of the right hand revealed significant main effects of *digit measure*, $F(2.06, 1570.16) = 12{,}235.56$, $p < 0.001$, $\eta^2_p = 0.94$, and *sex*, $F(1761) = 19.48$, $p < 0.001$, $\eta^2_p = 0.03$, and a significant *digit measure*sex* interaction, $F(2.06, 1570.16) = 7.44$, $p = 0.001$, $\eta^2_p = 0.01$. Significant main effects of *digit measure*, $F(2.00, 1520.49) = 12{,}345.43$, $p < 0.001$, $\eta^2_p = 0.94$, and *sex*, $F(1761) = 27.21$, $p < 0.001$, $\eta^2_p = 0.04$, and a significant *digit measure*sex* interaction, $F(2.00, 1520.49) = 15.07$, $p < 0.001$, $\eta^2_p = 0.02$, could also be observed for the left hand. Post-hoc *t*-tests revealed significant differences in every digit ratio between boys and girls (all $p \leq 0.004$), except for right, left, and aver-

		Boys			Girls			Comparison			Effect sizes and 95% CI	95% CI		
		M	*SD*	*N*	*M*	*SD*	*N*	*df*	*t*	*p*		Cohens *d*	lower	upper
2D	right	247.43	14.31	399	239.70	13.64	364	761	7.61	<.001		-0.55	-0.70	-0.41
	left	244.50	14.23	399	237.35	13.46	364	761	7.11	<.001		-0.52	-0.66	-0.37
	averaged	245.96	13.50	399	238.53	12.87	364	761	7.77	<.001		-0.56	-0.71	-0.42
3D	right	276.85	15.71	399	268.17	14.42	364	761	7.93	<.001		-0.58	-0.72	-0.43
	left	274.55	15.19	399	266.70	14.65	364	761	7.24	<.001		-0.53	-0.67	-0.38
	averaged	275.70	14.80	399	267.44	13.91	364	761	7.92	<.001		-0.57	-0.72	-0.43
4D	right	263.05	15.66	399	253.29	14.35	364	761	8.89	<.001		-0.64	-0.79	-0.50
	left	261.41	15.62	399	253.05	14.96	364	761	7.54	<.001		-0.55	-0.69	-0.40
	averaged	262.19	14.86	399	253.17	13.86	364	761	8.65	<.001		-0.64	-0.78	-0.49
5D	right	213.52	14.77	399	203.52	13.65	364	761	9.68	<.001		-0.70	-0.85	-0.56
	left	214.10	14.35	399	204.19	14.08	364	761	9.61	<.001		-0.70	-0.84	-0.55
	averaged	213.81	13.67	399	203.86	12.89	364	761	10.32	<.001		-0.75	-0.90	-0.60

Table 2. Means, standard deviations, post-hoc *t*-tests, and effect sizes for sex differences in digit lengths of both hands (N = 763).

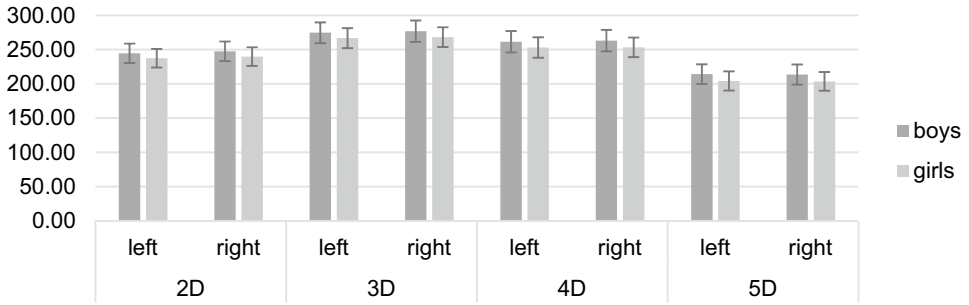


Figure 2. Means and standard deviations in right, left, and averaged digit lengths separately for boys and girls.

aged 2D:4D as well as left 3D:4D, where no significant sex difference could be observed (see Table 3). In general, girls had larger digit ratios than boys. The means and standard deviations are presented in Fig. 3.

Relative digit lengths. The 2 × 2 × 4 mixed ANOVA revealed a significant main effect of the factor *digit measure*, F (2.25, 1710.31) = 16,600.99, $p < 0.001$, $\eta^2_p = 0.96$, on relative digit lengths and significant *digit measure*sex*, F (2.25, 1710.31) = 18.57, $p < 0.001$, $\eta^2_p = 0.02$, and *hand*digit measure* interactions, F (2.24, 1702.63) = 26.96, $p < 0.001$, $\eta^2_p = 0.03$. The *hand*digit measure*sex* interaction was non-significant ($p = 0.474$). For the 2 × 4 mixed ANOVA separated by hand, a significant main effect of *digit measure*, F (2.19, 1665.92) = 12,085.74, $p < 0.001$, $\eta^2_p = 0.94$, and a significant *digit*sex* interaction, F (2.19, 1665.92) = 11.70, $p < 0.001$, $\eta^2_p = 0.02$, could be observed for the right hand. For the left hand, there was also a significant main effect of *digit measure*, F (2.30, 1750.94) = 11,999.04, $p < 0.001$, $\eta^2_p = 0.94$, and a significant *digit*sex* interaction, F (2.30, 1750.94) = 17.33, $p < 0.001$, $\eta^2_p = 0.02$. Post-hoc *t*-tests unveiled significantly larger rel2 (all $p \leq 0.020$) and rel3 (all $p \leq 0.004$) for girls compared to boys. rel4 did not significantly differ between girls and boys (all $p \geq 0.266$). In rel5, boys exhibited significantly larger relative digit lengths than girls (all $p < 0.001$; see Table 4). The means and standard deviations are presented in Fig. 4.

Right-left asymmetry in 2D:4D. The independent samples *t*-test showed no significant difference between girls' ($M = 0.01$, $SD = 0.04$) and boys' $D_{r\text{-}l}$ ($M = 0.01$, $SD = 0.04$), t (761) = 0.88, $p = 0.379$, $d = 0.06$.

Hand width, length, and index. A 2 × 2 mixed ANOVA with hand width as the outcome variable revealed significant main effects of the factors *hand*, F (1, 178) = 16.42, $p < 0.001$, $\eta^2_p = 0.08$, and *sex*, F (1, 178) = 15.34, $p < 0.001$, $\eta^2_p = 0.08$. The interaction between these factors was non-significant ($p = 0.186$). Post-hoc *t*-tests revealed significant differences between boys and girls in the right, left and average over both hands' width, with boys having wider hands than girls (all $p \leq 0.001$; see Table 5). The means and standard deviations are presented in Fig. 5.

For hand length, a 2 × 2 mixed ANOVA revealed a significant main effect of *sex*, F (1, 178) = 9.91, $p = 0.002$, $\eta^2_p = 0.05$. The *hand*sex* interaction was marginally significant, F (1, 99) = 2.90, $p = 0.091$, $\eta^2_p = 0.02$, while the main effect of *hand* was non-significant ($p = 0.475$). Post-hoc *t*-tests revealed significant sex differences in right, left and average over both hands' lengths, with boys having longer hands than girls (all $p \leq 0.001$; see Table 5). The means and standard deviations are presented in Fig. 5.

		Boys			Girls			Comparison			Effect sizes and 95% CI	95% CI		
		M	*SD*	*N*	*M*	*SD*	*N*	*df*	*t*	*p*		Cohens *d*	lower	upper
2D:4D	right	0.942	0.040	399	0.947	0.038	364	761	1.87	.061		0.14	-0.01	0.28
	left	0.936	0.038	399	0.939	0.038	364	761	0.98	.329		0.07	-0.07	0.21
	averaged	0.939	0.033	399	0.943	0.033	364	761	1.69	.092		0.12	-0.02	0.27
2D:5D	right	1.162	0.066	399	1.180	0.065	364	761	3.96	<.001		0.29	0.14	0.43
	left	1.144	0.059	399	1.164	0.060	364	761	4.78	<.001		0.35	0.20	0.49
	averaged	1.152	0.054	399	1.172	0.052	364	761	5.11	<.001		0.37	0.23	0.51
3D:4D	right	1.053	0.029	399	1.059	0.028	364	761	2.90	.004		0.21	0.07	0.35
	left	1.051	0.027	399	1.055	0.028	364	761	1.88	.060		0.14	-0.01	0.28
	averaged	1.052	0.023	399	1.057	0.023	364	761	2.87	.004		0.21	0.07	0.35
3D:5D	right	1.299	0.065	399	1.320	0.063	364	761	4.55	<.001		0.33	0.19	0.47
	left	1.285	0.060	399	1.309	0.061	364	761	5.52	<.001		0.40	0.26	0.54
	averaged	1.291	0.051	399	1.314	0.053	364	761	5.75	<.001		0.42	0.27	0.56
4D:5D	right	1.233	0.051	399	1.246	0.051	364	761	3.50	<.001		0.25	0.11	0.40
	left	1.223	0.049	399	1.241	0.050	364	761	5.15	<.001		0.37	0.23	0.52
	averaged	1.228	0.045	399	1.243	0.044	364	761	4.92	<.001		0.36	0.22	0.50

Table 3. Means, standard deviations, post-hoc *t*-tests, and effect sizes for sex differences in digit ratios of both hands ($N = 763$).

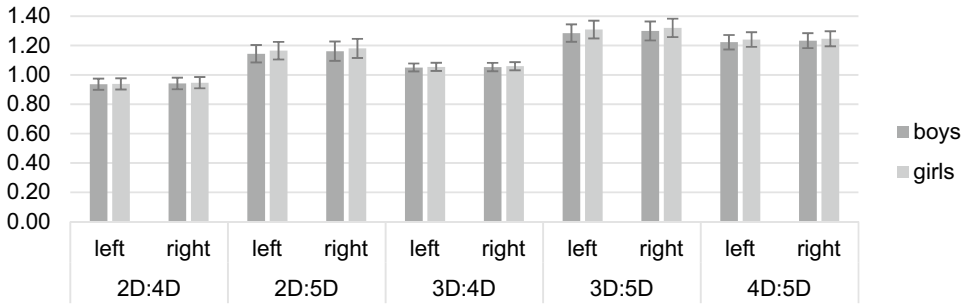


Figure 3. Means and standard deviations in right, left, and averaged digit ratios separately for boys and girls.

A 2 × 2 mixed ANOVA revealed a significant main effect of *hand*, $F(1, 178) = 17.76$, $p < 0.001$, $\eta^2_p = 0.09$, on the hand index. The main effect of *sex* and the *hand*sex* interaction were non-significant (both $p \geq 0.145$). Post-hoc *t*-tests showed no significant differences between boys' and girls' hand index (all $p \geq 0.134$; see Table 5). The means and standard deviations are presented in Fig. 6.

Discussion

Hand and digit measures are widely used to discriminate between the sexes and are hypothesized to reflect a different prenatal androgen exposure. In fact, there are robust sex differences in various measures, namely digit lengths and ratios, relative digit lengths as well as the hand width, length, and the hand index. However, those findings mostly rely on adult cohorts whereby findings in younger, specifically prepubertal cohorts remain more heterogeneous. Moreover, different measurement techniques impede the comparability of different studies as they result in considerable differences concerning measurement precision and possible bias. In the current study, we aimed to bridge the gap concerning the research of hand and digit measures as markers for prenatal androgen action in prepubertal cohorts by analyzing sex differences in a sample of 6-month-old infants and to give valuable methodological implications. Therefore, we analyzed the inter-examiner reliability as well as the test-retest reliability of digit length, hand width and length as well as the computed measures. Furthermore, we compared hand and digit measures which have been introduced in the literature, namely length of 2D, 3D, 4D, 5D, digit ratios 2D:4D, 2D:5D, 3D:4D, 3D:5D, 4D:5D, relative digit length rel2, rel3, rel4, rel5, directional asymmetry $D_{r\text{-}l}$ of 2D:4D, as well as hand width, length, and the hand index within one study using the same measurement technique, i.e. indirect measurements and a computer program.In the current study, the reliability analysis revealed excellent inter-examiner reliability while the test-retest reliability was only moderate to high. The results of this research provide supporting evidence for sex differences in digit lengths, ratios, and other digit and hand measures. Boys generally exhibited larger digits and bigger hands (i.e. hand width and length) with moderate to high effect sizes and smaller digit ratios compared to girls, a pattern that was evident for both right and left hands. Unexpectedly, the most commonly evaluated digit ratio, 2D:4D, showed no significant sex difference for left 2D:4D and only a marginally significant difference for the right and average of both hands

		Boys			Girls			Comparison			Effect sizes and 95% CI	95% CI		
		M	*SD*	*N*	*M*	*SD*	*N*	*df*	*t*	*p*		Cohens *d*	lower	upper
Rel2	right	0.247	0.007	399	0.249	0.007	364	761	2.50	.013		0.18	0.04	0.32
	left	0.246	0.006	399	0.247	0.006	364	761	2.33	.020		0.17	0.03	0.31
	averaged	0.247	0.006	399	0.248	0.006	364	761	2.87	.004		0.20	0.06	0.34
Rel3	right	0.277	0.005	399	0.278	0.005	364	761	4.03	<.001		0.28	0.14	0.42
	left	0.276	0.005	399	0.278	0.005	364	761	4.13	<.001		0.30	0.16	0.44
	averaged	0.276	0.004	399	0.278	0.004	364	761	4.76	<.001		0.33	0.18	0.47
Rel4	right	0.263	0.005	399	0.263	0.005	364	761	0.54	.588		-0.04	-0.18	0.10
	left	0.263	0.005	399	0.263	0.005	364	761	1.11	.266		0.08	-0.06	0.22
	averaged	0.263	0.004	399	0.263	0.004	364	761	0.33	.739		0.02	-0.12	0.17
Rel5	right	0.213	0.008	399	0.211	0.007	364	761	4.43	<.001		-0.32	-0.47	-0.18
	left	0.215	0.007	399	0.212	0.007	364	761	5.63	<.001		-0.41	-0.55	-0.27
	averaged	0.214	0.007	399	0.212	0.006	364	761	5.78	<.001		-0.43	-0.57	-0.29

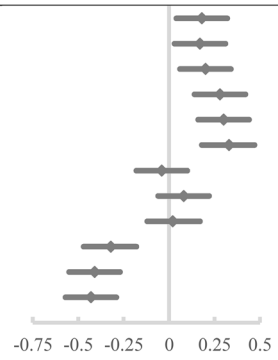


Table 4. Means, standard deviations, post-hoc *t*-tests, and effect sizes for sex differences in relative digit lengths of both hands (N = 763).

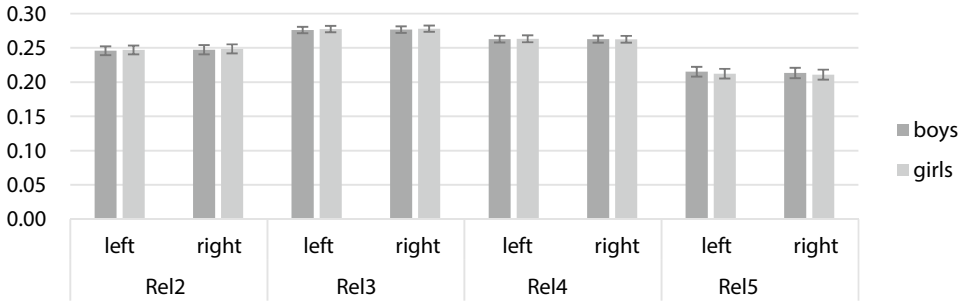


Figure 4. Means and standard deviations in right, left, and averaged relative digit lengths separately for boys, and girls.

with low effect sizes. For the relative digit lengths, low to moderate effect sizes and a different pattern of sex differences was observed for rel2 and rel3, with boys exhibiting smaller relative digit lengths than girls, whereas for rel5, they had larger relative digit lengths. There was no sex difference and low to moderate effects sizes in rel4, directional asymmetry of right and left 2D:4D (D_{r-l}) and the computed hand index.The reliability analysis confirmed that the inter-examiner reliability of digit and hand measures is nearly perfect, which is in line with other studies investigating indirect measurement techniques[27,40] and indicates high reliability and repeatability of those indirect and computer-assisted measurement techniques. Regarding the test-retest reliability of different digit and hand measures, the results ranged from high to moderate. Only a few studies controlled multiple measurements among different examiners and found generally good test-retest reliabilities in adult cohorts[27,42]. Mikac, Buško, Sommer and Hildebrandt (2016) note that the reliability of repeated hand scans can be controlled by accurate instructions and a standardized measurement[27], specifically by controlling the pressure with which hands are pressed onto the scanner glass, as this may lead to an error due to a shifting of important landmarks (e.g. flexion creases)[27]. These indications may not entirely apply to studies with young children because specifically very young children like infants cannot be adequately instructed and the examiner has to perform the hand scan. A further distortion by uncontrolled movements of the child and varying pressure applied during the scanning process may occur. Furthermore, because of the relatively higher amount of soft tissue in infants compared to adult cohorts[49], varying pressure may lead to an even greater displacement of important landmarks. Regarding studies that have investigated hand and digit measures in very young cohorts, only one other study applied an indirect measurement technique using hand scans and also found generally low to moderate effect sizes[16]. It becomes apparent that in general more direct measurement techniques have been applied in very young cohorts (see supplementary table 1). Given the test-retest reliability in the current study, it may be indeed more applicable to directly measure hands and digits in infants, e.g. by using calipers or radiographs, as the indirect measurement of digit lengths and measures may account for solid differences in measurements. However, due to the lack of studies investigating different measurement techniques within a sample, a comparison of different measurement techniques in infants is highly indicated. Furthermore, indirect measurements may be easier to implement compared to e.g. radiographs, and it becomes even more important to compare different measurement techniques within one study specifically in prepubertal, very young cohorts as well as to investigate the validity and reliability. Future studies should additionally investigate the test-retest reliability of measurement

		Boys			Girls			Comparison			Effect sizes and 95% CI	95% CI		
		M	*SD*	*N*	*M*	*SD*	*N*	*df*	*t*	*p*		Cohens *d*	lower	upper
Width	right	361.18	21.22	101	348.88	21.35	79	178	3.85	<.001		-0.58	-0.88	-0.28
	left	364.46	18.52	101	355.36	18.19	79	178	3.30	.001		-0.50	-0.79	-0.20
	averaged	362.82	17.68	101	352.12	18.82	79	178	3.92	<.001		-0.59	-0.89	-0.29
Length	right	632.52	29.10	101	616.80	31.62	79	178	3.46	.001		-0.52	-0.82	-0.22
	left	628.85	28.37	101	618.30	29.56	79	178	2.43	.016		-0.37	-0.66	-0.07
	averaged	630.69	27.26	101	617.55	28.43	79	178	3.15	.002		-0.47	-0.77	-0.18
Index	right	57.13	2.80	101	56.60	2.75	79	178	1.28	.203		-0.19	-0.49	0.10
	left	58.01	2.85	101	57.52	2.46	79	178	1.22	.223		-0.18	-0.48	0.11
	averaged	57.57	2.48	101	57.04	2.14	79	178	1.51	.134		-0.23	-0.52	0.07

Table 5. Means, standard deviations, post-hoc *t*-tests, and effect sizes for sex differences in hand width, length and index of both hands ($N = 180$).

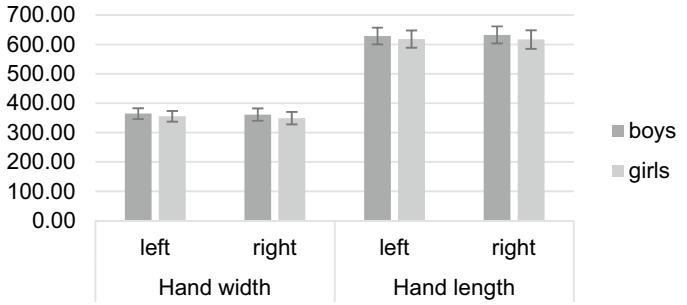


Figure 5. Means and standard deviations in right, left, and averaged hand width and length separately for boys and girls.

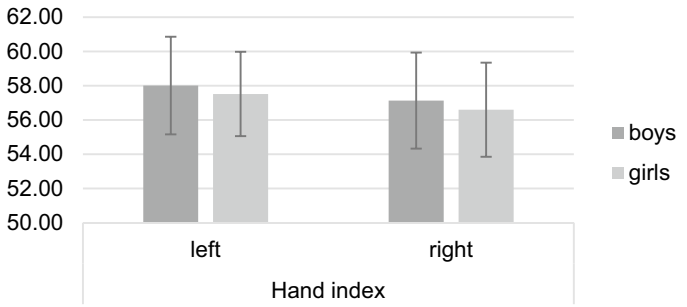


Figure 6. Means and standard deviations in right, left, and averaged hand index separately for boys and girls.

techniques and should not solely rely on the generally good inter- and intra-examiner reliabilities, especially in indirect measurements.

Regarding sex differences in different hand and digit measures, the results of the current study support the hypothesis that those measures can reliably differentiate between males and females as early as 6 months of age. The general assumption, that male hands and digits are larger than female whereby different digit ratios tend to be smaller in males compared to females, could also be supported[25,28]. However, these findings are contrasted by other studies, specifically in prepubertal cohorts, reporting considerable fluctuations in growth development as well as in digit measures[16,17,50]. Especially studies in prenatal or very young samples could not find sex differences in digit lengths and ratios from 9 to 40 weeks of gestational age[51] as well as in newborns born between 37 and 40 weeks of gestational age[52] or in hand length, width and index in deceased fetuses of 20 to 40 weeks of gestational age[38]. Studies investigating digit growth over a specific lifespan found larger digits in females from 2 years of age up to the age of 12 and a significant shift afterwards with males exhibiting longer digits after the age of 12[17,50]. This stands in direct contrast to our findings as we could replicate a male advantage in digit lengths that is reliably detected in adult cohorts. However, Gillam and colleagues (2008)[50] as well as Manning and Fink (2018)[17] did not examine the growth development between delivery and 2 years of age. The only available study reporting sex differences between 0 and 2 years of age, albeit just for 2D:4D, could find a significant difference

but solely at 2 weeks of age with small to moderate effect sizes[16]. They also report significant age effects and weak correlations between measurements taken at different time points[16]. This is in fact a comparable result to the moderate test-retest reliability in our study that has been discussed in the previous paragraph. Regarding the growth of the hand, a similar pattern emerges by viewing the available literature, where robust sex differences can be found in adult samples, while younger cohorts show no significant sex differences in hand length, width and index (see supplementary table 1). Furthermore, sex differences in younger, prepubertal cohorts are described as more fluctuating and in adult cohorts as robust[5,53]. However, it is important to note that hand measures are mostly considered in forensic contexts and adult samples[35–37], and in prepubertal cohorts most of the studies comparing hand length and width of girls and boys do that in terms of growth curves or prediction of body size in adult life[5,52,54]. There is a lack of studies considering hand length and width as markers for prenatal androgen exposure. In the current study, we included these measures as the influence of HOX genes and prenatal androgens, as stated in the introduction, refers to limb development and does not differentiate between the hand itself and individual digits[6–9]. The fact that malformations due to the expression of HOX genes can affect the hand itself as well as the digits, is another point that leads to the assumption of hand lengths and width as possible alternative markers[7,8]. Future research should investigate the biological basis of hand length and width regarding their ability as a reflection of the influence of HOX genes and prenatal androgens.

Considering digit ratios, the study of Knickmeyer and colleagues (2011) also shows no differences in 2D:4D in younger cohorts which is in line with our results[16]. This is in fact comparable to other findings that could not find sex differences in 2D:4D in younger cohorts[55–59]. This questions the assumption that 2D:4D is sexually dimorphic as early as the prenatal age[19,51] and independent of age effects[17]. With regard to the other digit ratios, our results suggest stronger sex differences in ratios using 5D as one of their components, which has been already promoted by Kumar and colleagues (2017), albeit their measurement technique significantly differed from ours[22]. There are other studies reporting sex differences in other digit ratios, however, a comparison to our results is limited as these studies investigated only pubertal or postpubertal cohorts[23–26,29]. Another study by Manning (2012) investigated alternative digit ratios in prepubertal cohorts and reported a sexual dimorphic pattern in 2D:3D, 2D:4D, 2D:5D and right hands' 3D:5D and 4D:5D, however, the author argues that these alternative ratios will not serve as good markers for sex differences since the age-dependent fluctuations are much more distinct compared to those in 2D:4D[12]. A general limitation of this assumption is the lack of longitudinal data[12] that also accounts for our findings. A study investigating digit ratios in a sample of 108-7 to 13-year-olds and the age-related changes after 4 years found significant sex differences in left and right 2D:4D, left 2D:3D, left 2D:5D and right 3D:5D, but only for the first measurement, and a significant sex difference in left 2D:5D that was also apparent for the second measurement. After correcting for multiple testing, only the sex difference in 2D:3D remained significant. However, contrary to the assumption of Manning (2012)[12], they have found a generally high stability for digit ratios between the two measurements. A study by McIntyre and colleagues (2005) analyzed the growth development between 1 month and 18 years of age and the sexual dimorphism of 2D, 3D, 4D as well as 2D:4D and 3D:4D and found significant differences in 3D:4D as early as the age of five[28]. Sex differences in 2D:4D occurred as early as 9 years of age. The authors conclude that digit ratios as markers for sex differences become more applicable with age, however, 3D:4D revealed sex differences in younger years more efficiently compared to 2D:4D[28]. There seems to be a general increase of sex differences in digit ratios with age and the literature supports the initial hypothesis that sex differences in prepubertal cohorts underlie significant age-related changes and thus findings remain more heterogeneous compared to adult respectively postpubertal cohorts.

As the discussed findings leave behind a considerable concern regarding the usefulness and applicability of hand and digit measures in prepubertal cohorts, it may be useful to mention the underlying hypothesis of these markers. It is proposed that the growth of hands and digits is influenced by prenatal androgen exposure. This could already be promoted by animal studies manipulating prenatal maternal and fetal hormone concentrations where a treatment with testosterone in pregnant Lewis rats led to a significant shortening of female and male offsprings' 2D as well as male offsprings' 4D[60]. The treatment had no effect on 2D:4D. In Wistar rats, a treatment with testosterone led to a shortening of the second digit and a simultaneous lengthening of the fourth digit in the offspring of pregnant rats compared to a control group that was treated with sesame oil[61]. In addition, the authors found a smaller 2D:4D in the testosterone group. In fact, Zheng & Cohn (2011) have shown that the growth of 2D and 4D in mice is controlled by the activity of androgen and estrogen receptors[62]. The authors showed that in general mice exhibit a similar pattern regarding 2D:4D with male mice having a smaller digit ratio than female mice. A deletion of the androgen receptor led to a larger 2D:4D ratio in males, while the deletion of the estrogen receptor led to a decrease of the 2D:4D ratio in males. Similarly, the inactivation of the androgen receptor led to a decrease in growth of the 4D while the inactivation of the estrogen receptor led to an increase. This growth pattern contributes to the sexual dimorphic pattern of the 2D:4D ratio in males and females. In general, a greater activity of androgen and estrogen receptors could be found in the 4D compared to the 2D[62]. A manipulation of hormonal concentrations during pregnancy is not possible in human studies due to ethical considerations. However, it is possible to compare the relative digit lengths in humans as proposed by Loehlin and colleagues (2009)[29] and, with regard to Suchonova et al. (2019)[60], Talarvičová et al. (2009)[61] and Zheng & Cohn (2011)[62], significant differences in the relative length of single digits are expected. In fact, we have found moderate to large differences in rel2 and rel3 with a female advantage, and a male advantage in rel5. Contrary to the assumption that 4D exhibits a greater activity of androgen and estrogen receptors[62], we found no differences in rel4, although 4D is hypothesized to be the most sensitive for the effects of androgens and estrogens. In sum, the relatively high heterogeneity in different studies considering prepubertal cohorts discussed in previous paragraphs suggests that the initial hypothesis should be reinvestigated as the high stability of sex differences in adult or postpubertal cohorts and the reported increase of sex differences with age suggests other factors than prenatal androgen exposure that contribute to the sexual dimorphic pattern of hand and digit measures. If

differences in digit lengths are primarily determined by the organizational influence of prenatal androgens and estrogens, a more stable sex difference in prepubertal cohorts should be expected.

It is well established that the human anatomy is lateralized and that specific bilateral structures are asymmetric[63]. This asymmetry in digits is hypothesized to rely on a different influence of prenatal androgens and can be measured as the directional asymmetry of right and left 2D:4D, $D_{r\text{-}l}$[32]. However, $D_{r\text{-}l}$ showed no significant difference between girls and boys of our sample, a finding that runs contrary to Manning et al. (2019)[64] but is in accordance with other studies that failed to reveal such a difference[33,34]. In the current study, we did not further investigate the differences between the left and right hand as the focus of the present study was on sex differences in different hand and digit measures. However, except for hand length, a significant main effect for the factor *hand* could be revealed. But this effect, albeit descriptively, did not promote the general assumption that sex differences are more pronounced in the right hand[11] as descriptive values were not always higher in the right hand (see Tables 2, 3, 4, and 5). A recent study showed that the asymmetry between the right and left hand in adults and children of 4 years of age is influenced by handedness, with right-handers exhibiting a more pronounced right-directional asymmetry compared to left-handers, thus emphasizing the influence of genetics on limb development[65]. In the current study, we did not control for handedness or other factors influencing the asymmetry and future studies should consider those alternative factors that may serve as explanations for differences in the right versus left hand. It may be additionally useful to compare $D_{r\text{-}l}$ in digit ratios other than 2D:4D as, in regard to the general assumption of a different influence of prenatal androgens on the right versus left side of the body, this should be analogously observable in other digit ratios. As our study aimed to evaluate hand and digit measures that have already been cited in the available literature, this was beyond the scope of the current study. However, this is an interesting approach for future studies.

Although our overall results promote the general assumption of sexually dimorphic hands and digits as well as different hand and digit measures, our study deals with several limitations that may be important in the context of sexually dimorphic anthropometric measurements in very young cohorts. As the initial study and the subsequent study design did not primarily investigate sex differences in hand and digit measures but mental rotation[43], the decision to examine a cohort of 6-month-olds was based on other theoretical considerations and hypotheses that did not focus on digit ratios as markers for prenatal androgen exposure. Furthermore, we acknowledge that while the results of the present sample regarding digit measures are based on a large sample of 763 infants, we could only measure hand width and length from 180 infants. Although a sample of 180 can be considered as a sufficient sample size, the missing data points need to be pointed out as a limitation of the study. Future studies should include in their study protocol a review of the hand scans so that non-evaluable scans can be identified early and rescanned which we unfortunately did not consider at the time the hand scans were taken. However, as the existing literature does not report many findings in cohorts between 0 and 2 years of age, our study gives important results to further compare different studies investigating sex differences in hand and digit ratios in very young cohorts. Furthermore, we could not control for the total body size as this variable was not evaluated. This may be additionally interesting as it would put the computed ratios in a more general context. Although we believe a strength of the present study is that we investigated the test-retest reliability, additional pre-scans were only available for a subset of our sample, as the initial procedure planned only the post-scans and the pre-scans were only implemented at the end of recruitment. Therefore, we had a limited number of available pre-scans. This was especially unfavorable for the test-retest reliability of hand width, length, and the hand index, as the few available pre-scans did not allow a meaningful analysis. Lastly, we did not correct for multiple testing. However, we reported the effect sizes and confidence intervals to promote the validity of our results.

In sum, our results provide supporting evidence for sex differences in different hand and digit measures as early as 6 months of age. However, the most prominent marker for the hypothesized relation between prenatal androgen exposure and the development of hands and digits, 2D:4D, did not reveal significant sex differences. Concerning the high heterogeneity in findings considering alternative digit ratios in prepubertal cohorts and the relatively high age-dependent fluctuation in digit growth comparing the more stable and generally larger sex differences in adult or postpubertal cohorts, there may be other factors that promote the sexually dimorphic pattern in hands and digits[66]. Kerrigan and Rogol (1992) argue, that the influence of sex hormones on growth development becomes more important with onset of puberty and that sex hormones have relevant, however complex and more secondary, moderating effects on the secretion of growth hormones[39], which supports the general impression of more homogeneous findings with increasing age. Nevertheless, specifically 2D:4D reveals significant and interesting correlations with several human behaviors known to differ between males and females, e.g. play behavior in children[67–69], as well as with several developmental disorders[70–72], psychiatric disorders[73] or even various types of cancer[74–79]. Similar results have been already reported for alternative digit ratios that could be linked to coronary heart diseases[80], attention deficit hyperactivity disorder[81], and externalizing and internalizing behaviors in children[82]. Although the underlying mechanisms, factors, and complex relations are not yet fully clarified, a valid and reliable measurement of hand and digit measures would be especially valuable in younger cohorts. Specifically for the early detection of disorders and diseases, it should be a major goal for future studies to further evaluate and provide for the quality of the hypothesized markers, i.e. hand and digit ratios. The current study provided important and valuable methodological implications and could support the assumptions of sex differences in different hand and digit measures as early as 6 months of age.

Data availability

The datasets analyzed in the current study are available in the Open Science Framework repository, https://osf.io/upzcx/?view_only=bf491c87e0ca4cab8ee62c7b0a841caf.

Received: 12 January 2021; Accepted: 23 April 2021
Published online: 26 May 2021

References

1. Arnold, A. P. & McCarthy, M. M. in *Neuroscience in the 21st Century* (ed N. D. Volkow D. W. Pfaff) Ch. 69, 2139–2168 (Springer Science + Business Media, 2016).
2. Hines, M. *Brain gender*. (Oxford University Press, 2004).
3. Eggers, S. & Sinclair, A. Mammalian sex determination—insights from humans and mice. *Chromosome Res.* **20**, 215–238. https://doi.org/10.1007/s10577-012-9274-3 (2012).
4. Amirsheybani, H. R. *et al.* The natural history of the growth of the hand: I Hand area as a percentage of body surface area. *Plastic Reconstruct. Surgery* **107**, 726–733. https://doi.org/10.1097/00006534-200103000-00012 (2001).
5. Matzdorff, I. Metrische und allometrische Wachstumsuntersuchungen an der menschlichen Hand. *Z. Morphol. Anthropol.* **59**, 158–184 (1967).
6. Manning, J., Scutt, D., Wilson, J. & Lewis-Jones, D. 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. https://doi.org/10.1093/humrep/13.11.3000 (1998).
7. Goodman, F. R. Limb malformations and the human HOX genes. *Am. J. Med. Genet.* **112**, 256–265. https://doi.org/10.1002/ajmg.10776 (2002).
8. Kondo, T., Zákány, J., Innis, J. W. & Duboule, D. Of fingers, toes, and penises. Nature **390**, 29. https://doi.org/10.1038/36234 (1997).
9. McIntyre, M. H. The use of digit ratios as markers for perinatal androgen action. *Reproduct. Biol. Endocrinol. RB&E* **4**, 10. https://doi.org/10.1186/1477-7827-4-10 (2006).
10. Hönekopp, J., Bartholdt, L., Beier, L. & Liebert, A. Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: New data and a meta-analytic review. *Psychoneuroendocrinology* **32**, 313–321. https://doi.org/10.1016/j.psyneuen.2007.01.007 (2007).
11. Hönekopp, J. & Watson, S. Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand. *Am. J. Hum. Biol.* **22**, 619–630. https://doi.org/10.1002/ajhb.21054 (2010).
12. Manning, J. T. in *Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease* (ed Victor R. Preedy) 841–851 (Springer New York, 2012).
13. Cohen-Bendahan, C. C., van de Beek, C. & Berenbaum, S. A. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neurosci. Biobehav. Rev.* **29**, 353–384. https://doi.org/10.1016/j.neubiorev.2004.11.004 (2005).
14. Arnold, A. P. & Breedlove, S. M. O. Organizational and activational effects of sex steroids on brain and behavior: a reanalysis. *Horm. Behav.* **19**, 469–498. https://doi.org/10.1016/0018-506X(85)90042-X (1985).
15. Arnold, A. P. The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of ammalian tissues. *Horm. Behav.* **55**, 570–578. https://doi.org/10.1016/j.yhbeh.2009.03.011 (2009).
16. Knickmeyer, R. C., Woolson, S., Hamer, R. M., Konneker, T. & Gilmore, J. H. 2D:4D ratios in the first 2 years of life: Stability and relation to testosterone exposure and sensitivity. *Horm. Behav.* **60**, 256–263. https://doi.org/10.1016/j.yhbeh.2011.05.009 (2011).
17. Manning, J. T. & Fink, B. Sexual dimorphism in the ontogeny of second (2D) and fourth (4D) digit lengths, and digit ratio (2D:4D). *Am. J. Hum. Biol.* **30**, e23138. https://doi.org/10.1002/ajhb.23138 (2018).
18. Breedlove, S. M. Minireview: Organizational hypothesis: instances of the fingerpost. *Endocrinology* **151**, 4116–4122. https://doi.org/10.1210/en.2010-0041 (2010).
19. Galis, F., Ten Broek, C. M., Van Dongen, S. & Wijnaendts, L. C. Sexual dimorphism in the prenatal digit ratio (2D:4D). *Arch. Sex Behav.* **39**, 57–62. https://doi.org/10.1007/s10508-009-9485-7 (2010).
20. Wong, W. I. & Hines, M. Interpreting digit ratio (2D:4D)-behavior correlations: 2D:4D sex difference, stability, and behavioral correlates and their replicability in young children. *Horm. Behav.* **78**, 86–94. https://doi.org/10.1016/j.yhbeh.2015.10.022 (2016).
21. McIntyre, M. H., Cohn, B. A. & Ellison, P. T. Sex dimorphism in digital formulae of children. *Am. J. Phys. Anthropol.* **129**, 143–150. https://doi.org/10.1002/ajpa.20240 (2006).
22. Kumar, S., Voracek, M. & Singh, M. Sexual dimorphism in digit ratios derived from dorsal digit length among adults and children. *Front. Endocrinol.* **8**, 41. https://doi.org/10.3389/fendo.2017.00041 (2017).
23. Kyriakidis, I. Data regarding 2D:4D and other digit ratios in Greek population. *Data Brief* **34**, 106724. https://doi.org/10.1016/j.dib.2021.106724 (2021).
24. Kyriakidis, I. & Papaioannidou, P. Epidemiologic study of the sexually dimorphic second to fourth digit ratio (2D:4D) and other finger ratios in Greek population. *Coll. Antropol.* **32**, 1093–1098 (2008).
25. McFadden, D. & Shubel, E. Relative lengths of fingers and toes in human males and females. *Horm. Behav.* **42**, 492–500. https://doi.org/10.1006/hbeh.2002.1833 (2002).
26. Dressler, S. G. & Voracek, M. No association between two candidate markers of prenatal sex hormones: digit ratios (2D:4D and other) and finger-ridge counts. *Dev. Psychobiol.* **53**, 69–78. https://doi.org/10.1002/dev.20488 (2011).
27. Mikac, U., Buško, V., Sommer, W. & Hildebrandt, A. Analysis of different sources of measurement error in determining second-to-fourth digit ratio, a potential indicator of perinatal sex hormones exposure. *Annu. Rev. Psychol.* **23**, 39–49. https://doi.org/10.21465/rp0023.0004 (2016).
28. McIntyre, M. H., Ellison, P. T., Lieberman, D. E., Demerath, E. & Towne, B. The development of sex differences in digital formula from infancy in the Fels Longitudinal Study. *Proc. R Soc. Lond. B Biol. Sci.* **272**, 1473–1479. https://doi.org/10.1098/rspb.2005.3100 (2005).
29. Loehlin, J. C., Medland, S. E. & Martin, N. G. Relative finger lengths, sex differences, and psychological traits. *Arch Sex Behav.* **38**, 298–305. https://doi.org/10.1007/s10508-007-9303-z (2009).
30. Stenstrom, E., Saad, G., Nepomuceno, M. V. & Mendenhall, Z. Testosterone and domain-specific risk: Digit ratios (2D:4D and rel2) as predictors of recreational, financial, and social risk-taking behaviors. *Pers. Individ. Dif.* **51**, 412–416. https://doi.org/10.1016/j.paid.2010.07.003 (2011).
31. Manning, J., Kilduff, L., Cook, C., Crewther, B. & Fink, B. Digit ratio (2D:4D): a biomarker for prenatal sex steroids and adult sex steroids in challenge situations. *Front Endocrinol. Lausanne* **5**, 9. https://doi.org/10.3389/fendo.2014.00009 (2014).
32. Manning, J. T. *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. (Rutgers University Press, 2002).
33. Coyne, S. M., Manning, J. T., Ringer, L. & Bailey, L. Directional asymmetry (right–left differences) in digit ratio (2D:4D) predict indirect aggression in women. *Pers. Individ. Dif.* **43**, 865–872. https://doi.org/10.1016/j.paid.2007.02.010 (2007).
34. Richards, G., Gomes, M. & Ventura, T. Testosterone measured from amniotic fluid and maternal plasma shows no significant association with directional asymmetry in newborn digit ratio (2D:4D). *J. Dev. Orig. Health Dis.* **10**, 362–367. https://doi.org/10.1017/S2040174418000752 (2019).
35. Kanchan, T. & Krishan, K. Anthropometry of hand in sex determination of dismembered remains—A review of literature. *J. Forensic Leg. Med.* **18**, 14–17. https://doi.org/10.1016/j.jflm.2010.11.013 (2011).
36. Krishan, K., Kanchan, T. & Sharma, A. Sex determination from hand and foot dimensions in a North Indian population. *J. Forensic Sci.* **56**, 453–459. https://doi.org/10.1111/j.1556-4029.2010.01652.x (2011).
37. Aboul-Hagag, K. E., Mohamed, S. A., Hilal, M. A. & Mohamed, E. A. Determination of sex from hand dimensions and index/ring finger length ratio in Upper Egyptians. *Egypt J. Forensic. Sci.* **1**, 80–86. https://doi.org/10.1016/j.ejfs.2011.03.001 (2011).

38. Raziye, D., Ceren, U., Kadir, D., Osman, S. & Mehmed, A. M. A radiological investigation on the hand development in human fetuses throughout the fetal period and an evaluation performed in terms of its clinical importance hand development. *Int. J. Morphol.* **34**, 1539–1552 (2016).
39. Kerrigan, J. R. & Rogol, A. D. The impact of gonadal steroid hormone action on growth hormone secretion during childhood and adolescence. *Endocr. Rev.* **13**, 281–298. https://doi.org/10.1210/edrv-13-2-281 (1992).
40. Kemper, C. J. & Schwerdtfeger, A. Comparing indirect methods of digit ratio (2D:4D) measurement. *Am. J. Hum. Biol.* **21**, 188–191. https://doi.org/10.1002/ajhb.20843 (2009).
41. Richards, G. & James, T. K. A comparison of self-measured and researcher-measured digit ratio (2D:4D). *Ann. Hum. Biol.* **46**, 527–530. https://doi.org/10.1080/03014460.2019.1674380 (2019).
42. Allaway, H. C., Bloski, T. G., Pierson, R. A. & Lujan, M. E. Digit ratios (2D:4D) determined by computer-assisted analysis are more reliable than those using physical measurements, photocopies, and printed scans. *Am. J. Hum. Biol.* **21**, 365–370. https://doi.org/10.1002/ajhb.20892 (2009).
43. Kaaz, T. & Heil, M. Infants' looking times in a 2-D mental rotation task. *Infant Child Dev.* **29**, e2167. https://doi.org/10.1002/icd.2167 (2020).
44. Autometric Software Measurement of 2D: 4D Ratios (http://www.facelab.org/debruine/Programs/autometric.php, 2004).
45. IBM SPSS Statistics for Windows, Version 27.0 v. 27.0 (IBM Corporation, Armonk, New York, 2020).
46. Ribeiro, E., Neave, N., Morais, R. N. & Manning, J. T. Direct versus indirect measurement of digit ratio (2D:4D): a critical review of the literature and new data. *Evolut. Psychol. Int. J. Evolut. Approaches Psychol. Behav.* **14**, 1474704916632536. https://doi.org/10.1177/1474704916632536 (2016).
47. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012 (2016).
48. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2 edn, (Lawrence Erlbaum Associates, 1988).
49. Kuzawa, C. W. Adipose tissue in human infancy and childhood: An evolutionary perspective. *Am. J. Phys. Anthropol.* **107**, 177–209. https://doi.org/10.1002/(SICI)1096-8644(1998)107:27+%3c177::AID-AJPA7%3e3.0.CO;2-B (1998).
50. Gillam, L., McDonald, R., Ebling, F. J. P. & Mayhew, T. M. Human 2D (index) and 4D (ring) finger lengths and ratios: cross-sectional data on linear growth patterns, sexual dimorphism and lateral asymmetry from 4 to 60 years of age. *J. Anat.* **213**, 325–335. https://doi.org/10.1111/j.1469-7580.2008.00940.x (2008).
51. Malas, M. A., Dogan, S., Evcil, E. H. & Desdicioglu, K. Fetal development of the hand, digits and digit ratio (2D:4D). *Early Hum. Dev.* **82**, 469–475. https://doi.org/10.1016/j.earlhumdev.2005.12.002 (2006).
52. Brandt, J. M., Allen, G. A., Haynes, J. L. & Butler, M. G. Normative standards and comparison of anthropometric data of white and black newborn infants. *Dysmorphol. Clin. Gen.* **4**, 121–137 (1990).
53. Jee, S.-C., Bahn, S. & Yun, M. H. Determination of sex from various hand dimensions of Koreans. *Forensic Sci. Int.* **257**(521), e521. https://doi.org/10.1016/j.forsciint.2015.10.014 (2015).
54. Marshall, W. A. Growth and sexual maturation in normal puberty. *Clin. Endocrinol. Metab.* **4**, 3–25. https://doi.org/10.1016/S0300-595X(75)80032-6 (1975).
55. Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R. & Manning, J. T. 2nd. to 4th digit ratios, fetal testosterone and estradiol. *Early Hum Dev* **77**, 23–28. https://doi.org/10.1016/j.earlhumdev.2003.12.002 (2004).
56. Alexander, G. M. & Saenz, J. Early androgens, activity levels and toy choices of children in the second year of life. *Horm. Behav.* **62**, 500–504. https://doi.org/10.1016/j.yhbeh.2012.08.008 (2012).
57. Buck, J. J., Williams, R. M., Hughes, I. A. & Acerini, C. L. In-utero androgen exposure and 2nd to 4th digit length radio. *Hum. Reprod.* **18**, 976–979. https://doi.org/10.1093/humrep/deg198 (2003).
58. Hwang, H. *et al.* The second-to-fourth digit ratio in cryptorchidism: a case-control study. *Korean J. Urol.* **55**, 140–144. https://doi.org/10.4111/kju.2014.55.2.140 (2014).
59. Barrett, E. *et al.* Digit ratio, a proposed marker of the prenatal hormone environment, is not associated with prenatal sex steroids, anogenital distance, or gender-typed play behavior in preschool age children. *J. Dev. Orig Health Dis.* **1**, 10. https://doi.org/10.1017/S2040174420001270 (2020).
60. Suchonova, M. *et al.* Does the 2nd and 4th digit ratio reflect prenatal androgen exposure?. *Bratisl. Lek. Listy* **120**, 703–710. https://doi.org/10.4149/bll_2019_119 (2019).
61. Talarovičová, A., Kršková, L. & Blažeková, J. Testosterone enhancement during pregnancy influences the 2D:4D ratio and open field motor activity of rat siblings in adulthood. *Horm. Behav.* **55**, 235–239. https://doi.org/10.1016/j.yhbeh.2008.10.010 (2009).
62. Zheng, Z. & Cohn, M. J. Developmental basis of sexually dimorphic digit ratios. *Proc. Natl. Acad. Sci. USA* **108**, 16289–16294. https://doi.org/10.1073/pnas.1108312108 (2011).
63. Manning, J. T., Fink, B., Neave, N. & Szwed, A. The second to fourth digit ratio and asymmetry. *Ann. Hum. Biol.* **33**, 480–492. https://doi.org/10.1080/03014460600802551 (2006).
64. Manning, J., Cook, C. & Crewther, B. Digit ratio (2D:4D) and testosterone supplementation. *Early Hum. Dev.* **139**, 104843. https://doi.org/10.1016/j.earlhumdev.2019.104843 (2019).
65. Kumar, S., Voracek, M. & Singh, M. The effects of hand preference and sex on right-left asymmetry in dorsal digit lengths among adults and children. *Early Human Dev.* **153**, 105293. https://doi.org/10.1016/j.earlhumdev.2020.105293 (2021).
66. Swift-Gallant, A., Johnson, B. A., Di Rita, V. & Breedlove, S. M. Through a glass, darkly: Human digit ratios reflect prenatal androgens, imperfectly. *Horm Behav.* **120**, 104686. https://doi.org/10.1016/j.yhbeh.2020.104686 (2020).
67. Hönekopp, J. & Thierfelder, C. Relationships between digit ratio (2D:4D) and sex-typed play behavior in pre-school children. *Pers. Ind. Dif* **47**, 706–710. https://doi.org/10.1016/j.paid.2009.06.007 (2009).
68. Körner, L. M., Schaper, M. L., Pause, B. M. & Heil, M. Parent-reports of sex-typed play preference in preschool children: relationships to 2d:4d digit ratio and older siblings' sex. *Arch. Sex. Behav.* **49**, 2715–2724. https://doi.org/10.1007/s10508-020-01662-6 (2020).
69. Mitsui, T. *et al.* The relationship between the second-to-fourth digit ratio and behavioral sexual dimorphism in school-aged children. *PLoS ONE* **11**, e0146849. https://doi.org/10.1371/journal.pone.0146849 (2016).
70. Hönekopp, J. Digit ratio 2D:4D in relation to autism spectrum disorders, empathizing, and systemizing: a quantitative review. *Autism Res. Official J. Int. Soc. Autism Res.* **5**, 221–230. https://doi.org/10.1002/aur.1230 (2012).
71. Schieve, L. A. *et al.* Associations between the 2nd to 4th digit ratio and autism spectrum disorder in population-based samples of boys and girls: findings from the study to explore early development. *J. Autism Dev. Disord.* **48**, 2379–2395. https://doi.org/10.1007/s10803-018-3495-z (2018).
72. Fink, B., Manning, J. T., Williams, J. H. G. & Podmore-Nappin, C. The 2nd to 4th digit ratio and developmental psychopathology in school-aged children. *Pers Individ. Dif.* **42**, 369–379. https://doi.org/10.1016/j.paid.2006.07.018 (2007).
73. Han, Y. *et al.* Association between the 2D:4D ratio and schizophrenia. *J. Int. Med. Res.* **48**, 300060520929148. https://doi.org/10.1177/0300060520929148 (2020).
74. Bunevicius, A. The association of digit ratio (2d : 4d) with cancer: a systematic review and meta-analysis. *Dis. Markers* **2018**, 7698193. https://doi.org/10.1155/2018/7698193 (2018).
75. Hong, L. *et al.* Digit ratio (2D:4D) in Chinese women with breast cancer. *Am. J. Hum. Biol.* **26**, 562–564. https://doi.org/10.1002/ajhb.22546 (2014).

76. Kasielska-Trojan, A. *et al.* Digit ratio (2D:4D) in women and men with lung cancer. *Sci. Rep.* **10**, 11369. https://doi.org/10.1038/s41598-020-68239-0 (2020).
77. Mendes, P. H. *et al.* Comparison of digit ratio (2D:4D) between Brazilian men with and without prostate cancer. *Prostate Cancer Prostatic. Dis.* **19**, 107–110. https://doi.org/10.1038/pcan.2015.62 (2016).
78. Sheng, Y. *et al.* Decreased digit ratio (2D:4D) and gastric cancer in Chinese men. *Early Hum. Dev.* **103**, 109–112. https://doi.org/10.1016/j.earlhumdev.2016.08.004 (2016).
79. Wang, L., Lu, H., Li, K.-K., Bai, C.-Y. & Ma, Z.-B. Digit ratio (2D:4D) in Chinese women with gastric cancer. *Am. J. Hum. Biol.* **30**, e23109. https://doi.org/10.1002/ajhb.23109 (2018).
80. Wang, L. *et al.* Digit ratio (2D:4D) and coronary artery disease in north Chinese women. *Early Hum Dev* **116**, 64–67. https://doi.org/10.1016/j.earlhumdev.2017.11.007 (2017).
81. Stevenson, J. C. *et al.* Attention deficit/hyperactivity disorder (ADHD) symptoms and digit ratios in a college sample. *Am J Hum Biol* **19**, 41–50. https://doi.org/10.1002/ajhb.20571 (2007).
82. Bruin, E., de Nijs, P. & Verhulst, H. A high 3D:5D ratio: A possible correlate of externalizing and internalizing problems: an exploratory study. *Europ. J. Psychiatry* **25**, 46–53. https://doi.org/10.4321/S0213-61632011000100006 (2011).

Author contributions

L.E., L.M.K., M.H., and N.K.S. contributed to the study idea and design. L.E. performed the primary analysis, and L.M.K., M.H. , and N.K.S. verified the analytical methods. L.E. designed the figures and tables and drafted a first version of the manuscript. L.M.K., M.H., G.R., and N.K.S. provided critique to refine the manuscript. All authors approved the final version of this manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-89590-w.

Correspondence and requests for materials should be addressed to L.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Investigating the reliability and sex differences of digit lengths, ratios, and hand measures in infants

Luisa Ernsten, Lisa M. Körner, Martin Heil, Gareth Richards, & Nora K. Schaal

Supplementary table 1 Summary and overview of the literature (in alphabetical order) underlying our research question and highlighting the available hand and digit measures. No claim is made to completeness.

Study	N (f/m)	Age	Measurement technique	Investigated hand/ digit measure	Sex differences	Effect size for sex difference
Aboul-Hagag et al., 2011	250 females, 250 males	≥ 18 years	direct, anthropometer, sliding caliper, ventral	hand length	m > f	1.45[a, b, f]-1.47[a, b, e]
				hand width	m > f	2.42[a, b, f]-2.45[a, b, e]
				hand index	m > f	1.49[a, b, e]-1.50[a, b, f]
				2D	m > f	1.64[a, b, e]-1.70[a, b, f]
				4D	m > f	2.03[a, b, e]-2.09[a, b, f]
				2D:4D	m < f	1.46[a, b, e]-1.62[a, b, f]
Dressler & Voracek, 2011	75 females, 75 males	mean age 25.9 years (*SD* = 9.2)	indirect, scans, digital Vernier caliper, ventral	indirect:	indirect:	indirect:
				2D:3D	n.s.	0.04[b, e]-0.09[b, f]
				2D:4D	m < f	0.39[b, f]-0.40[b, e]
				2D:5D	m < f	0.57[b, e]-0.68[b, f]
				3D:4D	m < f	0.42[b, f]-0.56[b, e]
				3D:5D	m < f	0.63[b, e]-0.68[b, f]
				4D:5D	m < f	0.43[b, e]-0.54[b, f]
			direct, digital Vernier caliper, ventral	direct:	direct:	direct:
				2D:3D	n.s.	0.04[b, f]-0.20[b, e]
				2D:4D	n.s.	0.18[b, e]-0.24[b, f]
				2D:5D	m < f	0.33[b, e]-0.55[b, f]
				3D:4D	m < f	0.38[b, f]-0.47[b, e]
				3D:5D	m < f	0.52[b, e]-0.60[b, f]
				4D:5D	m < f[f]	0.26[b, e]-0.48[b, f]
Galis et al., 2010	158 females, 169 males	14-28 weeks of gestational age	indirect, radiography, ruler, ventral	2D:4D	m < f[g]	0.23[a, c]

Knickmeyer et al., 2011	183 females, 181 males	0-2 years	indirect, photocopies, Vernier calipers, ventral	2D:4D	m < f [g]	0.13-0.33
Krishan et al., 2011	100 females, 100 males	17-20 years	direct, sliding caliper, ventral and dorsal	hand length hand width hand index	m > f m > f m > f	1.69[a, b, f]-1.72[a, b, e] 2.05[a, b, f]-2.10[a, b, e] 0.43[a, b, f]-0.50[a, b, e]
Kumar et al., 2017	Study I: 53 females, 51 males	Study I: 16-48 years	direct, Vernier calipers, dorsal	Study I: 2D:3D 2D:4D 2D:5D 3D:4D 3D:5D 4D:5D	Study I: n.s. n.s. m < f n.s. m < f m < f	Study I: 0.06[b, f]-0.15[b, e] 0.00[b, f]-0.31[b, e] 0.36[b, e]-0.72[b, f] 0.06[b, f]-0.28[b, e] 0.48[b, e]-0.69[b, f] 0.68[b, e]-0.83[b, f]
	Study II: 68 females, 86 males	Study II: 17-28		Study II: 2D:3D 2D:4D 2D:5D 3D:4D 3D:5D 4D:5D	Study II: n.s. n.s. m < f n.s. m < f m < f	Study II: 0.06[b, d, f]-0.18[b, e] 0.00[b, f]-0.15[b, e] 0.42[b, e]-0.48[b, d] 0.00[b, f]-0.05[b, e] 0.37[b, e]-0.50[b, f] 0.38[b, e]-0.57[b, f]
	Study III: 33 females, 31 males	Study III: 3-7.6 years		Study III: 2D:3D 2D:4D 2D:5D 3D:4D 3D:5D 4D:5D	Study III: n.s. n.s. n.s. n.s. m < f m < f	Study III: 0.33[b, e, f]-0.39[b, d] 0.07[b, f]-0.16[b, e] 0.25[b, e]-0.53[b, f] 0.13[b, e]-0.29[b, d, f] 0.44[b, e]-0.70[b, f] 0.46[b, e]-0.63[b, d]
Kyriakidis, 2021	80 females, 80 males	19-91 years	direct, electronic Vernier caliper, ventral	2D 3D 4D 5D 2D:3D 2D:4D 2D:5D 3D:4D 3D:5D	m > f m > f m > f m > f m < f m < f m < f m < f [f] n.s.	

				4D:5D	m < f [e]	
Kyriakidis & Papaioannidou, 2008	60 females, 60 males	19-25 years	direct, electronic Vernier caliper, ventral	2D	m > f	1.23[a, b, f]-1.24[a, b, e]
				3D	m > f	1.56[a, b, e]-1.67[a, b, f]
				4D	m > f	1.53[a, b, f]-1.60[a, b, e]
				5D	m > f	1.38[a, b, f]-1.57[a, b, e]
				2D:3D	m < f	0.62[a, b, e]-0.67[a, b, f]
				2D:4D	m < f	0.63[a, b, f]-0.75[a, b, e]
				2D:5D	m < f	0.62[a, b, f]-0.72[a, b, e]
				3D:4D	m < f [e]	0.16[a, b, f]-0.36[a, b, e]
				3D:5D	m < f [e]	0.16[a, b, f]-0.57[a, b, e]
				4D:5D	m < f [e]	0.09[a, b, f]-0.35[a, b, e]
Loehlin et al., 2009	400 females, 300 males	2/3 10-16 years, 1/3 ≥17 years	indirect, scans, computer program	rel2	m < f	0.46[b, f]-0.66[b, e]
				rel3	m < f	0.09[b, f]-0.22[b, e]
				rel4	m > f	0.10[b, f]-0.18[b, e]
				rel5	m > f	0.29[b, f]-0.33[b, e]
				2D:3D	m < f	0.32[b, f]-0.42[b, e]
				2D:4D	m < f	0.36[b, f]-0.58[b, e]
				2D:5D	m < f	0.39[b, f]-0.50[b, e]
				3D:4D	m < f	0.14[b, f]-0.33[b, e]
				3D:5D	m < f	0.22[b, f]-0.27[b, e]
				4D.5D	m < f	0.13[b, e]-0.17[b, f]
Manning, 2002 (Manning et al., 2000)	661 females, 567 males	children and adults	indirect, scans, Vernier calipers, ventral direct, Vernier calipers, ventral	$D_{r\text{-}l}$	n.s.	0.04[a, b]
Manning, 2012	340 females, 340 males	2-18 years	direct, Vernier calipers, ventral	2D:3D	m < f [g]	0.17[a, c, e]-0.30[a, c, f]
				2D:4D	m < f [g]	0.26[a, c, e]-0.30[a, c, f]
				2D:5D	m < f [g]	0.32[a, c, f]-0.35[a, c, e]
				3D:4D	n.s. [g]	0.03[a, c, f]- 0.15[a, c, e]
				3D:5D	m < f [e, g]	0.14[a, c, f]-0.27[a, c, e]
				4D:5D	m < f [e, g]	0.11[a, c, f]-0.17[a, c, e]
Manning & Fink, 2018	Sample I: 340 females, 340 males (40 per age group)	Sample I: 2-18 years	Sample I: direct, Vernier calipers, ventral	Sample I: 2D	Sample I: m > f [g]	Sample I: 0.20[a, c, e]
				4D	m > f [g]	0.38[a, c, e]

				2D:4D	m < f	0.26[a, c, e]
	Sample II: 42,843 females, 46,402 males	Sample II: 18-30 years	Sample II: self-measurement, ruler, ventral	Sample II: 2D 4D 2D:4D	Sample II: m > f [g] m > f [g] m < f	Sample II: 0.45[a, c, e] 0.49[a, c, e] 0.21[a, c, e]
McIntyre et al., 2005	56 females, 68 males	1 month-18 years	indirect, radiography, computer program	2D 3D 4D 2D:4D 3D:4D	m > f m > f m > f m < f [g] m < f [g]	 0.35[a, c, f] 0.44[a, c, f]
McIntyre et al., 2006	537 females, 523 males	2-10 years	indirect, radiography, computer program	2D:4D 3D:4D	m < f m < f	0.24[a, b, f] 0.61[a, b, f]
McFadden & Shubel, 2002	62 females, 60 males	heterosexual males *M* =19.0, heterosexual females *M* =19.1, homosexual males *M* =22.0, homosexual females *M* =20.7 years	indirect, scans, computer program	2D:3D 2D:4D 2D:5D 3D:4D 3D:5D 4D:5D	m < f m < f m < f m < f m < f n.s.	0.49[b, f]-0.50[b, e] 0.74[b, f]-0.85[b, e] 0.68[b, f]-0.76[b, e] 0.54[b, f]-0.72[b, e] 0.39[b, f]-0.51[b, e] 0.13[b, e, f]
Raziye et al., 2016	27 females, 23 males	20-40 weeks of gestational age	indirect, mammography and radiography, digital caliper	2D 3D 4D 5D 2D:4D hand length hand width hand index	n.s. n.s n.s n.s n.s n.s n.s. n.s.	
Richards et al., 2019	54 females, 52 males	*median* = 27.18 h after birth	indirect, scans, Vernier calipers, ventral	2D:4D $D_{r\text{-}l}$	m < f [f] n.s.	0.21[a, b, e] – 0.57[a, b, f] 0.30[a, b]
Stenstrom et al., 2011	235 females, 178 males	17-44 years	direct, digital calipers, ventral	2D:4D rel2	m < f m < f	0.30[b, e] 0.31[b, e]

Wong & Hines, 2016	70 females, 56 males	three age groups: 20-26 months, 27-33 months, 34-40 months	indirect, scans, ventral, in pixel using a computer program	2D:4D	m < f	0.61[c]

Note: n.s. – not significant, [a] Computed from test statistic, [b] Comparison of group means, [c] Main effect of factor sex, [d] Mean of right and left hand, [e] right hand, [f] left hand, [g] age effects.

Supplementary table 2 Intra-class correlations and 95% confident intervals of digit length measurements of two independent examiners.

			ICC	95% CI	
				upper	lower
2D	pre	right	.97	.96	.98
		left	.96	.95	.97
	post	right	.96	.95	.97
		left	.95	.95	.96
3D	pre	right	.97	.96	.98
		left	.97	.96	.98
	post	right	.97	.96	.97
		left	.95	.95	.96
4D	pre	right	.96	.95	.98
		left	.97	.97	.98
	post	right	.96	.96	.97
		left	.94	.93	.95
5D	pre	right	.97	.95	.98
		left	.95	.93	.97
	post	right	.96	.95	.96
		left	.95	.94	.95

Supplementary table 3 Intra-class correlations and 95% confident intervals of hand length and width measurements of two independent examiners.

			ICC	95% CI	
				upper	lower
Hand length	pre	right	.96	.93	.97
		left	.94	.89	.96
	post	right	.95	.93	.96
		left	.93	.92	.95
Hand width	pre	right	.86	.78	.92
		left	.96	.92	.97
	post	right	.94	.93	.95
		left	.95	.94	.96

Supplementary table 4 Test-retest reliability of pre and post scans for boys, girls, and overall.

		Boys			Girls			Overall		
		r	*p*	*N*	*r*	*p*	*N*	*r*	*p*	*N*
2D	right	.77	<.001	66	.86	<.001	64	.81	<.001	130
	left	.78	<.001	66	.84	<.001	64	.81	<.001	130
	averaged	.84	<.001	66	.92	<.001	64	.88	<.001	130
3D	right	.81	<.001	66	.90	<.001	64	.85	<.001	130
	left	.83	<.001	66	.87	<.001	64	.85	<.001	130
	averaged	.88	<.001	66	.94	<.001	64	.91	<.001	130
4D	right	.80	<.001	66	.72	<.001	64	.77	<.001	130
	left	.88	<.001	66	.84	<.001	64	.87	<.001	130
	averaged	.91	<.001	66	.86	<.001	64	.89	<.001	130
5D	right	.68	<.001	66	.74	<.001	64	.72	<.001	130
	left	.86	<.001	66	.82	<.001	64	.85	<.001	130
	averaged	.84	<.001	66	.88	<.001	64	.87	<.001	130
2D:4D	right	.58	<.001	66	.37	.002	64	.50	<.001	130
	left	.55	<.001	66	.68	<.001	64	.63	<.001	130
	averaged	.65	<.001	66	.68	<.001	64	.67	<.001	130
2D:5D	right	.42	<.001	66	.54	<.001	64	.49	<.001	130
	left	.62	<.001	66	.59	<.001	64	.63	<.001	130
	averaged	.62	<.001	66	.67	<.001	64	.67	<.001	130
3D:4D	right	.62	<.001	66	.37	.003	64	.52	<.001	130
	left	.56	<.001	66	.50	<.001	64	.53	<.001	130
	averaged	.71	<.001	66	.56	<.001	64	.66	<.001	130
3D:5D	right	.49	<.001	66	.65	<.001	64	.59	<.001	130
	left	.69	<.001	66	.62	<.001	64	.67	<.001	130
	averaged	.69	<.001	66	.75	<.001	64	.74	<.001	130
4D:5D	right	.54	<.001	66	.70	<.001	64	.64	<.001	130
	left	.68	<.001	66	.65	<.001	64	.67	<.001	130
	averaged	.69	<.001	66	.80	<.001	64	.75	<.001	130
rel2	right	.46	<.001	66	.43	<.001	64	.46	<.001	130
	left	.56	<.001	66	.63	<.001	64	.61	<.001	130
	averaged	.60	<.001	66	.65	<.001	64	.64	<.001	130
rel3	right	.56	<.001	66	.62	<.001	64	.60	<.001	130
	left	.66	<.001	66	.55	<.001	64	.61	<.001	130
	averaged	.73	<.001	66	.72	<.001	64	.73	<.001	130
rel4	right	.70	<.001	66	.42	<.001	64	.57	<.001	130
	left	.54	<.001	66	.66	<.001	64	.61	<.001	130
	averaged	.70	<.001	66	.70	<.001	64	.71	<.001	130
rel5	right	.45	<.001	66	.65	<.001	64	.56	<.001	130
	left	.69	<.001	66	.61	<.001	64	.67	<.001	130
	averaged	.67	<.001	66	.74	<.001	64	.72	<.001	130
$D_{r\text{-}l}$		.40	.001	66	.33	.008	64	.36	<.001	130

Study 3:

Ernsten, L., Körner, L. M., Heil, M., Schaal, N. K. (2024). The association between 2D:4D digit ratio and sex-typed play in children with and without siblings. *Scientific Reports 14*(1), 15231. https://doi.org/10.1038/s41598-024-65739-1

Author contributions: Ernsten, L.: Conceptualization; Data curation, Formal analysis, Methodology, Project administration, Visualization, Writing – original draft; Körner, L. M.: Conceptualization, Methodology, Writing – review & editing; Heil, M.: Conceptualization, Methodology, Investigation, Supervision, Writing – Review & Editing; Schaal, N. K.: Conceptualization, Methodology, Supervision, Writing – Review & Editing

Impact Factor (2022): 4.6 (Journal Citation Reports, 2023c)

scientific reports

Check for updates

OPEN

The association between 2D:4D digit ratio and sex-typed play in children with and without siblings

Luisa Ernsten[1]✉, Lisa M. Körner[1,2], Martin Heil[1] & Nora K. Schaal[1]

The 2D:4D digit ratio is commonly used as a surrogate possibly reflecting prenatal testosterone levels. Indirect evidence comes from studies investigating the association between 2D:4D and human characteristics that likely relate to prenatal testosterone. In children, sex-typed play reveals large sex differences early in development and an influence of prenatal testosterone is likely. Findings on the association between 2D:4D and children's sex-typed play are heterogeneous and other influences on the development of sex-typed play have been suggested, most of all social influences like siblings, their sex and birth order. The current study examined the association between right and left 2D:4D, a proposed surrogate for prenatal testosterone exposure, which was assessed in right and left hands of *N* = 505 6-month-old children, and sex-typed play behavior, which was evaluated 3.5 years later using the Pre-School Activities Inventory (PSAI), and the influence of siblings. To capture differential effects of siblings' sex and birth order, dummy-coded variables were used reflecting having no siblings as well as older or younger sisters or brothers. Multiple regression models were used to investigate the association between PSAI scores and sex, right and left 2D:4D, being a singleton as well as having an older or younger sister or brother. It was shown that sex and having an older brother were significant predictors for sex-typed play. Effects were further disentangled by conducting separate regression analyses in boys and girls. In boys, a significant association between PSAI scores and having an older brother was revealed, in girls, no significant associations were found. Results are discussed highlighting the non-significant association between 2D:4D and children's sex-typed play, which weakens the applicability of 2D:4D as a surrogate reflecting influences of prenatal T. Further, the importance of social factors like siblings on children's sex-typed play is discussed.

The 2D:4D digit ratio was introduced by Manning et al.[1] who found a significant sex difference (males < females) and a negative association of 2D:4D with sperm count. The authors suggested that 2D:4D is likely modified by prenatal intrauterine testosterone levels, relatively stable across human development, and therefore a suitable marker for prenatal testosterone exposure[1]. Since then, the sexually differentiated pattern was replicated in various studies, however, with only moderate effect sizes[2]. As approaches for more direct evidences on the association between 2D:4D and prenatal testosterone are limited due to ethical considerations in humans, indirect approaches have been conducted in clinical samples. Studies suggested a masculinized (i.e., smaller) 2D:4D in individuals with altered intrauterine testosterone concentrations like individuals with congenital adrenal hyperplasia (CAH) who are exposed to high levels of testosterone during gestation[3–6]. Likewise, feminized (i.e., larger) 2D:4D was found in individuals with complete androgen insensitivity syndrome (CAIS)[7,8], a clinical condition characterized by non-functioning androgen receptors. In these studies, effect sizes were generally small to medium and the variability of 2D:4D in clinical samples was comparable to healthy samples[3–8]. Other studies have focused on establishing associations between 2D:4D and other human characteristics as it is commonly assumed that androgens, especially testosterone, have organizational effects and shape brain connections already in-utero[9].

Different studies have found significant associations between 2D:4D and aggression (see, e.g.,[10–12], but see[13]), or athletic performances[14] in adults. During childhood, sex-typed play behavior gathered special interest in past

[1]Faculty of Mathematics and Natural Sciences, Institute of Experimental Psychology, Heinrich Heine University, Düsseldorf, Germany. [2]Department of General Pediatrics, Neonatology and Pediatric Cardiology, University Children's Hospital, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University, Düsseldorf, Germany. ✉email: Luisa.Ernsten@hhu.de

research. It shows considerably large effect sizes[15], emerges as early as 1 year of age[16], and a biological origin seems likely as it is also observable in non-human primates[17,18]. In human studies, it was shown that females with CAH show masculinized play behavior compared to unaffected females[19,20]. Further, individuals with CAIS show feminized behavior compared to non-affected controls[20]. Isolated gonadotropin-releasing-hormone deficiency (IGD), a condition with comparably low production of gonadal hormones during pregnancy, analogously was found to relate to less gender-conformity during childhood reported by affected adult males[21]. It is assumed that hormones like testosterone influence brain development as early as the fetal stage in such a way that higher levels of testosterone result in a masculinization, which can also affect behavior later in life[9]. Several studies investigated the direct association between sex-typed play and testosterone. Considering prenatal testosterone, different studies could not find a significant association of amniotic fluid testosterone levels with sex-typed play in toddlers[22] and pre-school children[23,24]. Otherwise, Auyeung et al.[25] found a significant positive relationship between maternal amniotic fluid testosterone concentrations and masculinized play behavior in on average 8.5-year-old boys and girls. Hines et al.[26] obtained testosterone in maternal blood samples during pregnancy and found significant positive associations with male-typed play behavior at age 3.5 years, however, only in girls. Other studies investigated postnatal testosterone levels. Alexander and Saenz[27] could not find a significant association between children's salivary testosterone and different aspects of children's play at age 3 to 4 months. Otherwise, Lamminmäki et al.[28] found a significant association between urinary testosterone that was obtained weekly in newborn infants until 6 months of age and sex-typed play behavior at 14 months of age in the hypothesized direction (i.e., higher urinary testosterone related to more masculinized play behavior). Both studies have assessed testosterone during a time-frame where male testosterone levels peak while female testosterone levels remain low shortly after birth and possibly also have organizational effects on the brain (a time-frame also called 'mini-puberty'[29]). In sum, study results promote the idea that (prenatal) testosterone relates to sex-typed behavior during childhood even when measured with different techniques and approaches.

In children, different studies have examined the relationship between 2D:4D and sex-typed play. Some studies found that smaller 2D:4D was associated with more masculinized play behavior, however, only in boys[30,31] or only in girls[32], whereas other studies found it in both sexes[33]. Another recent study found no significant association between 2D:4D and sex-typed play behavior[34]. The studies used the same tool for the assessment of sex-typed play behavior, the elaborated Pre-School Activities Inventory[35] (although Hönekopp and Thierfelder[30] used a different method for the calculation of the total PSAI score). Although results should be highly comparable due to the methodology, results on the association between 2D:4D and children's sex-typed play behavior remained heterogeneous.

With regard to differing study results, researchers suggested that controlling for other variables, like social influences, may influence the association between 2D:4D and sex-typed play behavior. Different models exist for the explanation of sex-typed behavior due to social influences (see, e.g.,[36]). Parents potentially have a crucial function as role models in children's development and may also directly affect children by reinforcing or punish certain behavior[37]. While there is some evidence that mothers and fathers treat their children differently, Morawska[38] found only limited evidence in their systematic review that children show different behavior due to gendered parenting. Further, social learning theories and social modelling, e.g., through siblings and peers, have been suggested to be important sources for the development of sex-typed behavior[39,40]. Research has found that siblings' sex and birth order may be two relevant factors that affect the sibling relationship[41]. Stoneman et al.[42] described differential behavior patterns of older and younger sisters and brothers in different sibling constellations. They found that same-sex siblings engaged in more gender-typical play, whereas other-sex siblings engaged in sex-typed play that was typical for the older sibling's sex[42]. Similar results have been reported by Rust et al.[43], who additionally investigated sex-typed play in singleton children. It was found that children with older same-sex siblings exhibited highest ratings of sex-congruent play behavior, whereas children with older opposite-sex siblings exhibited lowest ratings of sex-congruent play behavior[43]. Singleton children showed more sex-typed play behavior compared to children with opposite-sex siblings, however, less compared to children with same-sex siblings[43]. Less is known about the influence of younger siblings on older siblings' sex-typed play behavior. Hughes et al.[44] argued that younger siblings facilitate social skills in children, whereas Sang and Nelson[45] found differential effects of younger siblings on older siblings and emphasized the importance to distinguish between different sibling constellations.

Regarding the association between 2D:4D and sex-typed play, some of the aforementioned studies also investigated the number and sex of siblings as an easy to quantify measure for potential social influences. Hönekopp and Thierfelder[30] found no differences in sex-typed play behavior between children with brothers or sister. Mitsui et al.[31] found that the existence of brothers or sisters was associated with more masculinized or feminized play behavior, respectively. Lastly, Körner et al.[32] found that both 2D:4D and the existence of older siblings independently explained variance in sex-typed play, underscoring the importance to include both biological and social factors while investigating possible sources for sex-typed behavior. As the number and sex of siblings could be important for sex-typed behavior and is easily assessible, it seems reasonable to include this factor.

In sum, there is evidence that smaller 2D:4D, which presumably reflects more intrauterine testosterone exposure, can be associated with more masculinized play behavior in children, although, evidence differs substantially in the available literature. Further, it may be beneficial to investigate the influence of socializing factors, e.g., siblings, on sex-typed play. The aim of the current study was to investigate the association between 2D:4D and children's sex-typed play. As past research found diverging results, the potential influence of siblings in the context of socialization and its effect on sex-typed play will be additionally investigated. A significant association between 2D:4D and children's sex-typed play could be interpreted as a form of external validity for the applicability of 2D:4D as a marker for prenatal testosterone. However, as prior results substantially differ, it may be beneficious to include other variables that could affect this influence. Therefore, siblings are included as an additional factor. To capture diverging effects of siblings, different constellations of sibling's sex and birth order

will be incorporated. It is assumed in accordance to prior research that, specifically, older same-sex siblings relate to sex-congruent play behavior, whereas other-sex siblings possibly relate to sex-incongruent play behavior. As the influence of younger siblings has not been reviewed to a greater extent, this will be investigated in a more explorative way. Exceeding approaches of prior studies, the current study will also investigate the influence of being a singleton child on sex-typed play behavior, as research suggests that singleton children exhibit less sex-typed play behavior compared to children with same-sex siblings, however, more sex-typed play behavior compared to children with other-sex siblings.

Methods

Participants

As part of a larger study, German-speaking families with newborn children were recruited between 2013 and 2018 when infants were 6 months of age. Families were invited to the Institute of Experimental Psychology at Heinrich Heine University Düsseldorf, Germany, in order to take part in a mental rotation experiment (a subsample was published elsewhere, see[46]). In a follow-up 3.5 years later, parents were asked to rate their child's sex-typed behavior with the PSAI using an online questionnaire, whether children have siblings and whether and for how long they attended pre-school. Children were on average 6 months old ($M = 195.21$ days, $SD = 8.24$) when hand scans were taken and on average 3.5 to 4 years ($M = 46.25$ months, $SD = 3.74$) when their parents completed the online questionnaire. All of the children except for $n = 7$ attended pre-school by the time of the online questionnaire. The online questionnaire was completed by mothers in 89.33% of the cases, by fathers in 5.33% and by both parents in 5.33% of the cases. Almost all families and infants were White and from middle-class backgrounds. All parents and/or legal guardians gave their written informed consent for participation. The current study was approved by the local ethics committee of the Faculty of Mathematics and Natural Sciences of Heinrich Heine University Düsseldorf, Germany, and was in accordance with the Declaration of Helsinki.

Hand scans were taken from a total of 1391 children (705 boys and 683 girls), which were also used for another work of our group[47]. For the current study, only complete data sets with valid hand measures of second and fourth digits of right and left hands as well as a completed online questionnaire were used, leaving $n = 519$ cases. Additionally, $n = 5$ cases of right 2D:4D, $n = 4$ cases of left 2D:4D, and $n = 5$ cases of the PSAI score were excluded as the score ranged more than +/- 2 IQR from the median, leaving a final data set of $n = 505$ (240 girls, 265 boys). Of these, $n = 134$ (66 girls, 68 boys) were singletons.

Materials

Digit ratio (2D:4D)

Both right and left hands of 6-months-old children were scanned by an examiner using a FUJITSU fi-60F image scanner. The ventral surface of right and left hands was independently pushed softly onto the scanner glass and covered with a towel to increase contrast. For determining the length of digits, the freeware program *AutoMetric*[48] was used. For each digit, the length between the midpoint of the ventral proximal crease to the tip was measured. A 100 dpi monitor was used (100 pixels = 2.54 cm). The 2D:4D digit ratio was estimated by dividing the length of the second digit (2D) by the length of the fourth digit (4D). Indirect measurement techniques like hand scans can achieve larger sex differences and higher measurement precision than direct measurements[49]. Further, *Autometric* shows a high reliability for digit measurements and is superior to other computer-based measurement techniques[32]. Hand scans were rated by two independent raters, who were blind to the sex of the children. Intra-class correlations varied between 0.95 and 0.97, which can be seen as almost perfect. The two ratings were averaged for each digit on each hand to increase reliability.

Preschool activities inventory

The Preschool Activities Inventory (PSAI) is a standardized questionnaire by Golombok and Rust[35] used to measure sex-typed play behavior in children. In the present study, the German version described by Hönekopp and Thierfelder[30] was used. Sex-typed play behavior is measured via 24 items, divided into three groups: 7 toy items (e.g., 'guns', 'jewellery'), 11 activity items (e.g., 'fighting', 'playing house (e.g., cleaning, cooking)') and 6 character items (e.g., 'likes to explore new surroundings', 'likes pretty things'). Twelve items each are characterized as typically male and female. Parents were asked to rate the preferences of their child for the different toys and actions and their child's temperamental characteristics on a 5-point Likert-scale. A higher score indicates more typically male play behavior, a lower score more typically female play behavior[35]. The total PSAI score was calculated as follows according to Golombok and Rust[35]:

$$PSAIscore = 48.25 + 1.1(sumof\text{"male"}\ items - sumof\text{"female"}\ items) \quad (1)$$

Procedure

After arrival, written informed consent of caregivers was obtained. Next, the mental rotation task was performed[46]. Examiners then took hand scans of both right and left hands from children without giving further information about the interpretation of their childrens' finger length. Parents received a refund of their travel expenses. Families were contacted again approximately 3.5 years later. Parents followed an invitation via e-mail and answered a questionnaire using an open source online survey tool[50] and, again, gave their informed consent for participation, recording, and storage of data. Then, parents were asked to answer an online questionnaire including the PSAI[35]. Further, they stated who filled out the questionnaire, whether and for how long the child attended preschool, and whether and how many older and/or younger brothers and/or sisters the child had. The

study was approved by the local Ethics Committee of the Faculty of Mathematics and Natural Sciences of the Heinrich Heine University Düsseldorf, Germany.

Statistical analyses

To test whether 2D:4D and PSAI scores differed between the sexes, independent samples *t* tests were conducted.

To test for the influence of siblings on sex-typed play behavior, a multiple linear regression model with the criterion PSAI score and the predictors *sex*, *2D:4D* (right and left), *singletons* (dummy coded as 0 = no singleton, 1 = singleton), *older sisters* (dummy coded as 0 = no older sisters, 1 = at least one older sister), *younger sisters* (dummy coded as 0 = no younger sisters, 1 = at least one younger sister), *older brothers* (dummy coded as 0 = no older brothers, 1 = at least one older brother), and *younger brothers* (dummy coded as 0 = no younger brothers, 1 = at least one younger brother) was tested. Pearson correlations are reported for all study variables (PSAI score, sex, right and left 2D:4D, and dummy-coded variables singletons, older sisters, younger sisters, older brothers, younger brothers). To further disentangle effects of sex, two separate multiple linear regression models incorporating the same predictor variables (right and left 2D:4D, singletons, older sisters, younger sisters, older brothers, younger brothers) were run separately for subsamples of boys and girls.

Levels of significance were set to 0.05 for all comparisons. Pearson's *r* was interpreted as follows: small effect $r \geq 0.10$, medium effect $r \geq 0.30$, and large effect $r \geq 0.50$[51]. Cohen's *d* was interpreted as small effect $d \geq 0.20$, medium effect $d \geq 0.50$, and large effect $d \geq 0.80$[51].

Results

Right and left 2D:4D as well as PSAI scores were tested for normal distribution in boys and girls separately using Shapiro–Wilk tests and revealed no violation for normality assumption (all $p \geq 0.173$).

First, sex differences in 2D:4D and PSAI scores were tested for significance. Overall, girls had larger digit ratios and lower PSAI scores compared to boys. There were no significant sex differences in the overall sample in left 2D:4D, $t\,(503) = 0.92$, $p = 0.180$, $d = 0.08$, only right 2D:4D differed significantly between the sexes, $t\,(503) = 1.67$, $p = 0.048$, $d = 0.16$. PSAI scores differed significantly between girls and boys of the overall sample, $t\,(503) = 29.16$, $p < 0.001$, $d = 2.60$.

The association between sex-typed play behavior and siblings was further analyzed. Chi-square tests revealed that boys and girls did not significantly differ in the number of older and/or younger sisters and brothers (all $p > 0.173$, two-sided). Pearson correlations of the study variables are shown in Table 1. The multiple linear regression model with the predictors *sex*, *right 2D:4D*, *left 2D:4D*, *singletons*, *older sisters*, *younger sisters*, *older brothers*, and *younger brothers* was overall significant, $F\,(8, 496) = 112.03$, $p < 0.001$, with an $R^2 = 0.64$ (adjusted $R^2 = 0.64$). PSAI scores could be significantly predicted by the factors *sex*, and *older brothers (*see Table 2 for the results of the multiple linear regression model). To detect differential influences of siblings on PSAI scores of boys and girls, the multiple regression analysis was again tested separately for boys and girls using the predictors *right 2D:4D*, *left 2D:4D*, *singletons*, *older sisters*, *younger sisters*, *older brothers*, and *younger brothers*. In boys, the overall model was significant, $F\,(7, 257) = 3.54$, $p = .001$, with an $R^2 = .09$ (adjusted $R^2 = .06$). For boys, only having older brothers was a significant predictor of the PSAI score (see Table 3). For girls, the overall model did not reach significance, $F\,(7, 232) = 0.63$, $p = .734$, with an $R^2 = .02$ (adjusted $R^2 = -.01$). None of the predictors reached statistical significance (see Table 3).

A sensitivity analysis revealed that, given an α error probability of 0.05, a power of 0.95, the total sample size $N = 505$ and a total of eight predictors in the analysis, an effect size $f^2 = 0.05$ can be found with $\lambda = 23.10$ and a critical $F\,(8, 496) = 1.96$.

Discussion

The current study aimed to examine the relationship between 2D:4D digit ratio measured at 6 months of age and concurrent sex-typed play behavior of nearly 4-year-old girls and boys, as well as the influence of siblings by investigating the influence of having no siblings, older sisters, younger sister, older brothers and younger brothers. In the overall sample, only small sex differences could be found in 2D:4D (males < females) and this difference was only significant in right hands. The sex difference in PSAI scores, however, could be considered

Variables			1	2	3	4	5	6	7	8
1	PSAI		–							
2	Sex		− 0.79***	–						
3	2D:4D	right	− 0.07	0.07*	-					
4		left	− 0.06	0.04	0.42***	–				
5	Singletons		0.01	0.02	0.11**	− 0.01	–			
6	Sisters	older	− 0.05	0.03	0.03	− 0.00	− 0.29***	–		
7		younger	− 0.08*	0.03	− 0.11**	− 0.04	− 0.30***	− 0.16***	–	
8	Brothers	older	0.10*	0.00	0.00	− 0.01	− 0.33***	0.00	− 0.16***	–
9		younger	0.00	− 0.04	− 0.03	0.05	− 0.32***	− 0.17***	− 0.25***	− 0.16***

Table 1. Pearson correlations of study variables. $N = 505$; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ (two-tailed).

Variable					95% CI			
		B	*SE*	β	Lower	Upper	*t*	*p*
Intercept		73.61	13.15		47.77	99.44	5.60	< 0.001***
Sex		− 25.91	0.88	− 0.79	− 27.64	− 24.18	− 29.37	< 0.001***
2D:4D	right	− 5.79	12.74	− 0.01	− 30.81	19.23	− 0.46	0.650
	left	− 7.48	13.14	− 0.02	− 33.30	18.33	− 0.57	0.569
Singletons		2.11	1.88	0.06	− 1.59	5.80	1.12	0.264
Sisters	older	− 0.59	1.57	− 0.01	− 3.67	2.49	− 0.38	0.707
	younger	− 0.99	1.76	− 0.02	− 4.45	2.47	− 0.56	0.576
Brothers	older	4.40	1.50	0.11	1.46	7.35	2.94	0.003**
	younger	− 0.04	1.73	− 0.00	− 3.44	3.37	− 0.02	0.983

Table 2. Results of the multiple linear regression model testing the association between sex, 2D:4D, and siblings on PSAI scores. The variables singletons (0 = not a singleton, 1 = singleton) as well as *older sisters*, *younger sisters*, *older brothers*, and *younger brothers* were entered as dummy-coded variables (0 = no older/younger sister/brother, 1 = at least one older/younger sister/brother). *** $p < 0.001$, ** $p < 0.01$. $N = 505$;

Model	Variable				95% CI			
		B	*SE*	β	Lower	Upper	*t*	*p*
Boys								
Intercept		79.13	17.16		45.35	112.92	4.61	< 0.001***
2D:4D	Right	− 13.09	16.72	− 0.05	− 46.03	19.84	− 0.78	0.434
	Left	− 4.51	17.55	− 0.02	− 39.08	30.05	− 0.26	0.797
Singletons		1.53	2.56	0.07	− 3.50	6.56	0.60	0.550
Sisters	Older	− 3.05	2.22	− 0.12	− 7.41	1.32	− 1.38	0.170
	Younger	− 3.55	2.40	− 0.15	− 8.27	1.18	− 1.48	0.140
Brothers	Older	4.45	2.00	0.20	0.52	8.38	2.23	0.027*
	Younger	− 1.71	2.37	− 0.08	− 6.38	2.96	− 0.72	0.472
Girls								
Intercept		43.79	20.36		3.66	83.91	2.15	0.033*
2D:4D	Right	− 1.07	19.70	− 0.00	− 39.88	37.74	− 0.05	0.957
	Left	− 9.05	19.87	− 0.03	− 48.19	30.09	− 0.46	0.649
Singletons		2.19	2.82	0.09	− 3.36	7.73	0.78	0.438
Sisters	Older	1.25	2.25	0.05	− 3.19	5.68	0.56	0.580
	Younger	0.91	2.62	0.04	− 4.26	6.08	0.35	0.728
Brothers	Older	4.03	2.30	0.17	− 0.50	8.56	1.75	0.081
	Younger	1.06	2.57	0.04	− 4.01	6.13	0.41	0.680

Table 3. Results of separate multiple linear regression models testing the association between 2D:4D and siblings on PSAI scores for boys and girls. The variables singletons (0 = not a singleton, 1 = singleton) as well as *older sisters*, *younger sisters*, *older brothers*, and *younger brothers* were entered as dummy-coded variables (0 = no older/younger sister/brother, 1 = at least one older/younger sister/brother). ***$p < 0.001$, *$p < 0.05$. $n = 265$ boys and $n = 240$ girls.

large. As predicted, boys were reported to exhibit more masculinized play behavior, while girls were reported to show more feminized play behavior. The regression model in the overall sample was significant and revealed a significant influence of sex and that having an older brother related to significantly more male-typed play behavior in the current sample. The effect of older brothers on male-typed play behavior was also found in the subsample of boys, in girls, none of the predictors was significantly associated with PSAI scores.

Our analysis revealed only small or non-significant sex differences in 2D:4D, which contradicts the assumption of sex differences regardless of age and aligns with studies indicating small to non-existent sex differences in 2D:4D of very young cohorts[34,52,53]. Several studies have indicated that sex differences in 2D:4D become larger with age (see, e.g., Butovskaya et al.[54], Ernsten et al.[55], Trivers et al.[56]; but see also Lolli et al.[57], and Forstmeier[58], for the complexity of human allometry and how it can be reflected in simple ratios), therefore, it may not be surprising that we did not find sex differences in this young cohort of 6 months old children. It is commonly assumed that sex differences in 2D:4D are reliably detectable after the age of 2 years[1,59,60], whereas in younger cohorts only small or non-existent sex differences can be obtained[52]. In younger cohorts, some authors already suggested that measurement techniques may contribute to difficulties in finding larger effect sizes, as a higher amount of soft tissue in very young children possibly impairs the reliability of obtained hand scans[61]. Furthermore, standardized

measurements can be more difficult to perform in very young cohorts as adherence to instructions possibly lacks[47]. McIntyre et al.[62] found the digit ratio of the third to the fourth finger (3D:4D) to be a significantly better discriminator between boys and girls and argue that the measurement of longer digit segments possibly produces less measurement error compared to shorter digits. In a prior study, McIntyre et al.[53] found sex differences in both 2D:4D and 3D:4D, however, the sex difference in 3D:4D was evident much earlier compared to 2D:4D (i.e., 3D:4D at 5 years and 2D:4D at 9 years of age). Therefore, 2D:4D may not be the best marker to detect sex differences in very young cohorts. In sum, the cohort of the current study seemed to be in a critical age concerning the reliable detection of sex differences in 2D:4D. Nevertheless, although the current study did not find large sex differences, a descriptive difference could be found in the hypothesized direction (i.e., males < females) in our sample for both the right and left 2D:4D.

The sex difference in sex-typed play could be considered large, with significantly more feminized play behavior in nearly 4-year-old girls, and significantly more masculinized play behavior in nearly 4-year-old boys according to parents' reports. This aligns with the broad body of literature[9,15,30–34] and emphasizes the overall finding of large sex differences in sex-typed play behavior in pre-school children.

The question of interest of the current study, however, was the association between sex-typed play and 2D:4D as a surrogate of prenatal testosterone exposure and incorporating different sibling constellations as a possible social influence. 2D:4D and sex-typed play behavior did not relate in our sample. Also Barrett et al.[34] did not find any significant association between 2D:4D and sex-typed play, whereas Wong and Hines[33] found it in both sexes. However, both study groups did not investigate the influence of siblings. Hönekopp and Thierfelder[30] as well as Mitsui et al.[31] reported significant associations between 2D:4D and overall PSAI scores only in boys, with Hönekopp and Thierfelder[30] reporting no significant associations between siblings and PSAI scores, while Mitsui et al.[31] found that older same-sex siblings related to more sex-congruent play behavior in children. Both studies used correlational analyses between PSAI scores and sibling variables. Körner et al.[32] reported a significant association between 2D:4D and sex-typed play as well as that older brothers significantly related to more male-typed play behavior, however, only in girls. These authors also used multiple regression models incorporating both 2D:4D and sibling variables as predictors for PSAI scores. Körner et al. [32] argued that effects of variations of prenatal testosterone in a normal range may be evident only up to a specific threshold[32] and that in boys, who are exposed to generally higher levels of prenatal testosterone, this variation possibly has no further effect[26,32]. While our study results agree with the seemingly greater influence of older siblings, in the current sample, this was only true for having older brothers. Also Rust et al.[43] described that older brothers may reveal a larger effect on sex-typed play compared to older sister. They outlined that these differential effects may stem from cultural variations in how gender-normative or non-normative behavior is accepted in girls and boys[43]. A comparable result was found by Braun and Davidson[63], where boys who conformed to gender norms were most liked by their peers, whereas boys who did not conformed to gender norms were least liked. Conversely, girls who did not conform to gender norms were most liked by peers[63]. In sibling constellations, this may suggest that male siblings possibly act in a way that is more sex-congruent, whereas female siblings display a broad spectrum of behaviors and therefore become less sex-congruent.

In a more explorative approach, the current study included being a singleton and having younger siblings as additional variables that have not been investigated to a greater extent in other studies. However, these sibling constellations did not reveal any association with sex-typed play in children of the current sample. Rust et al.[43] discussed the differential influence of siblings on sex-typed behavior and argued that same-sex siblings and specifically older siblings shape sex-congruent behavior while children without siblings seem to be more variable in their sex-typed behavior, however, show more sex-congruent behavior compared to children with siblings of the opposite sex. The existence of younger siblings, on the other hand, possibly encourages nurturing and more caregiving behavior regardless of sex[44], a behavior that is typically characterized as feminine. It is further possible that children with younger siblings engage to a greater extent in sibling differentiation, which means to engage in sex-incongruent behavior to set oneself apart from the same-sex sibling[43]. Girls seemed to be less influenced in sex-typed play by siblings of either-sex in the current sample, and there are different studies on whether there is a difference in reinforcing or attenuating sex-typed behavior (e.g., through parents) between girls and boys (see, e.g.,[64,65]). A recent review, however, did not find robust evidence for differential reinforcement of sex-typed behavior between girls and boys[38].

In the current study, the statistical effect of older brothers exceeded the effect of 2D:4D on sex-typed play behavior, with 2D:4D having no significant effect on sex-typed play behavior. One could argue that the effect of social influences, like siblings, has prevailed the effect of biological factors as indicated by 2D:4D as a surrogate for prenatal testosterone on sex-typed play behavior in the current study. This would underscore the importance of social influences on pre-school children's sex-typed play behavior[66,67]. The effect size of 2D:4D was possibly too low to reliably differentiate between the sexes and to reveal a significant association with sex-typed play, especially when social influences like siblings were present. A larger sample could be indicated, although other studies found the hypothesized significant findings in a smaller sample compared to the current[32].

The current study had some limitations. First, the study population was relatively homogenous in terms of ethnicity and socio-economic status. Specifically as 2D:4D significantly varies with respect to ethnicity[68], results may not be generalizable. Next, the PSAI is a tool for the assessment of children's sex-typed play by reports of parents, therefore, a distortion by parents' perceptions of gender norms cannot be ruled out and, additionally, parents' perception of their child's play behavior may be influenced by siblings and subsequently by comparisons between children. Sex-typed play in children is a multifaceted construct and can be measured or observed in many ways. Other researches have used other tools to assess sex-typed play, like choices and preferences of playing tools[22,27], and also found associations with 2D:4D[27]. Also the PSAI itself captures sex-typed play on different subscales (toy, activity, and character items, see[35]), which are used to calculate an overall score that should represent children's sex-typed play. However, the PSAI possibly does not capture the multifaceted nature of sex-typed

play or, more generally spoken, of gender role behavior of children. Therefore, other measures for sex-typed play or gender role behavior and their association with 2D:4D should be investigated in future studies. The current study did not account for other social or cognitive influences that may impact the development of sex-typed play, like other same-sex or other-sex peers, teachers, other family members, or sex-role identities of the children[69], as well as for gender and/or sex diversity. The literature discusses non-binary and genderqueer sexual differentiation and identification that go beyond the scope of the current study. Nevertheless, sexual differentiation and identification other than the binary conception are important aspects to have in mind while discussing the source of sex/gender differentiation. A more general limitation could be the usefulness of 2D:4D as a marker for prenatal testosterone itself. Evidence in humans for direct associations between 2D:4D and prenatal testosterone is scarce and heterogeneous[55,70–72], although, some evidence exists and 2D:4D is a commonly used surrogate to investigate associations of prenatal testosterone with other variables of interest[73]. As effect sizes between 2D:4D and sources of prenatal testosterone have to be generally considered as small[73], there seems to be another source of variance that has not yet been identified. Also Hönekopp and Watson[2] noted the significance of identifying these other sources that may affect 2D:4D in one of the first meta analyses on 2D:4D and to control for them in future studies. Therefore, the results have to be interpreted with caution.

In sum, the results of the current study suggest to critically review the usefulness of 2D:4D as a proxy for prenatal T, as no significant associations were found between 2D:4D and sex-typed play in the current study. Further, the results of significant influences of older brothers on more male-typed play behavior further promote the importance of social influences on children's sex-typed behavior.

Data availability

The datasets analyzed in the current study are available in the Open Science Framework repository, https://osf.io/3cja7/.

Received: 8 December 2023; Accepted: 24 June 2024
Published online: 02 July 2024

References

1. Manning, J., Scutt, D., Wilson, J. & Lewis-Jones, D. 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. https://doi.org/10.1093/humrep/13.11.3000 (1998).
2. Hönekopp, J. & Watson, S. Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand. *Am. J. Hum. Biol.* **22**, 619–630. https://doi.org/10.1002/ajhb.21054 (2010).
3. Brown, W. M., Hines, M., Fane, B. A. & Breedlove, S. M. Masculinized finger length patterns in human males and females with congenital adrenal hyperplasia. *Horm. Behav.* **42**, 380–386. https://doi.org/10.1006/hbeh.2002.1830 (2002).
4. Ökten, A., Kalyoncu, M. & Yariş, N. The ratio of second- and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Early Hum. Dev.* **70**, 47–54. https://doi.org/10.1016/s0378-3782(02)00073-7 (2002).
5. Richards, G. *et al.* Digit ratio (2D:4D) and congenital adrenal hyperplasia (CAH): Systematic literature review and meta-analysis. *Horm. Behav.* https://doi.org/10.1016/j.yhbeh.2020.104867 (2020).
6. Buck, J. J., Williams, R. M., Hughes, I. A. & Acerini, C. L. In-utero androgen exposure and 2nd to 4th digit length radio. *Hum. Reprod.* **18**, 976–979. https://doi.org/10.1093/humrep/deg198 (2003).
7. Berenbaum, S. A., Bryk, K. K., Nowak, N., Quigley, C. A. & Moffat, S. Fingers as a marker of prenatal androgen exposure. *Endocrinology* **150**, 5119–5124. https://doi.org/10.1210/en.2009-0774 (2009).
8. van Hemmen, J., Cohen-Kettenis, P. T., Steensma, T. D., Veltman, D. J. & Bakker, J. Do sex differences in CEOAEs and 2D:4D ratios reflect androgen exposure? A study in women with complete androgen insensitivity syndrome. *Biol. Sex. Differ.* **8**, 11. https://doi.org/10.1186/s13293-017-0132-z (2017).
9. Hines, M. Neuroscience and sex/gender: Looking back and forward. *J. Neurosci.* **40**, 37–43. https://doi.org/10.1523/jneurosci.0750-19.2019 (2020).
10. Björkqvist, K. Gender differences in aggression. *Curr. Opin. Psychol.* **19**, 39–42. https://doi.org/10.1016/j.copsyc.2017.03.030 (2018).
11. Butovskaya, M., Burkova, V., Karelin, D. & Filatova, V. The association between 2D:4D ratio and aggression in children and adolescents: Cross-cultural and gender differences. *Early.Hum. Dev.* https://doi.org/10.1016/j.earlhumdev.2019.07.006 (2019).
12. Butovskaya, M. *et al.* Oxytocin receptor gene polymorphism (rs53576) and digit ratio associates with aggression: Comparison in seven ethnic groups. *J. Physiol. Anthropol.* **39**, 20. https://doi.org/10.1186/s40101-020-00232-y (2020).
13. Turanovic, J. J., Pratt, T. C. & Piquero, A. R. Exposure to fetal testosterone, aggression, and violent behavior: A meta-analysis of the 2D:4D digit ratio. *Aggress. Viol. Behav.* **33**, 51–61. https://doi.org/10.1016/j.avb.2017.01.008 (2017).
14. Hönekopp, J. & Schuster, M. A meta-analysis on 2D:4D and athletic prowess: Substantial relationships but neither hand out-predicts the other. *Personal. Individ. Differ.* **48**, 4–10. https://doi.org/10.1016/j.paid.2009.08.009 (2010).
15. Davis, J. T. M. & Hines, M. How large are gender differences in toy preferences? A systematic review and meta-analysis of toy preference research. *Arch. Sex. Behav.* **49**, 373–394. https://doi.org/10.1007/s10508-019-01624-7 (2020).
16. Todd, B. K. *et al.* Sex differences in children's toy preferences: A systematic review, meta-regression, and meta-analysis. *Infant Child Dev.* https://doi.org/10.1002/icd.2064 (2018).
17. Lonsdorf, E. V. Sex differences in nonhuman primate behavioral development. *J. Neurosci. Res.* **95**, 213–221. https://doi.org/10.1002/jnr.23862 (2017).
18. Marley, C. L., Pollard, T. M., Barton, R. A. & Street, S. E. A systematic review of sex differences in rough and tumble play across non-human mammals. *Behav. Ecol. Sociobiol.* **76**, 158. https://doi.org/10.1007/s00265-022-03260-z (2022).
19. Hines, M. Sex steroids and Human Behavior: Prenatal Androgen Exposure and Sex-typical Play Behavior in Children. *Ann. N. Y. Acad. Sci.* **1007**, 272–282. https://doi.org/10.1196/annals.1286.026 (2003).
20. Cohen-Bendahan, C. C., van de Beek, C. & Berenbaum, S. A. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neurosci. Biobehav. Rev.* **29**, 353–384. https://doi.org/10.1016/j.neubiorev.2004.11.004 (2005).
21. Shirazi, T. N. *et al.* Low perinatal androgens predict recalled childhood gender nonconformity in men. *Psychol. Sci.* **33**, 343–353. https://doi.org/10.1177/09567976211036075 (2022).
22. van de Beek, C., van Goozen, S. H., Buitelaar, J. K. & Cohen-Kettenis, P. T. Prenatal sex hormones (maternal and amniotic fluid) and gender-related play behavior in 13-month-old Infants. *Arch. Sex. Behav.* **38**, 6–15. https://doi.org/10.1007/s10508-007-9291-z (2009).

23. Spencer, D. *et al.* Prenatal androgen exposure and children's gender-typed behavior and toy and playmate preferences. *Hormon. Behav.* https://doi.org/10.1016/j.yhbeh.2020.104889 (2021).
24. Knickmeyer, R. C. *et al.* Gender-typed play and amniotic testosterone. *Dev. Psychol.* **41**, 517–528. https://doi.org/10.1037/0012-1649.41.3.517 (2005).
25. Auyeung, B. *et al.* Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. *Psychol. Sci.* **20**, 144–148. https://doi.org/10.1111/j.1467-9280.2009.02279.x (2009).
26. Hines, M. *et al.* Testosterone during Pregnancy and Gender Role Behavior of Preschool Children: A longitudinal population study. *Child Devel.* **73**, 1678–1687 (2002).
27. Alexander, G. M. & Saenz, J. Early androgens, activity levels and toy choices of children in the second year of life. *Horm. Behav.* **62**, 500–504. https://doi.org/10.1016/j.yhbeh.2012.08.008 (2012).
28. Lamminmäki, A. *et al.* Testosterone measured in infancy predicts subsequent sex-typed behavior in boys and in girls. *Horm. Behav.* **61**, 611–616. https://doi.org/10.1016/j.yhbeh.2012.02.013 (2012).
29. Hines, M. *et al.* The early postnatal period, mini-puberty, provides a window on the role of testosterone in human neurobehavioural development. *Curr. Opin. Neurobiol.* **38**, 69–73. https://doi.org/10.1016/j.conb.2016.02.008 (2016).
30. Hönekopp, J. & Thierfelder, C. Relationships between digit ratio (2D:4D) and sex-typed play behavior in pre-school children. *Pers. Individ. Dif.* **47**, 706–710. https://doi.org/10.1016/j.paid.2009.06.007 (2009).
31. Mitsui, T. *et al.* The Relationship between the Second-to-Fourth Digit Ratio and Behavioral Sexual Dimorphism in School-Aged Children. *PLoS One* https://doi.org/10.1371/journal.pone.0146849 (2016).
32. Körner, L. M., Schaper, M. L., Pause, B. M. & Heil, M. Parent-Reports of Sex-Typed Play Preference in Preschool Children: Relationships to 2D:4D Digit Ratio and Older Siblings' Sex. *Arch. Sex. Behav.* **49**, 2715–2724. https://doi.org/10.1007/s10508-020-01662-6 (2020).
33. Wong, W. I. & Hines, M. Interpreting digit ratio (2D:4D)-behavior correlations: 2D:4D sex difference, stability, and behavioral correlates and their replicability in young children. *Horm. Behav.* **78**, 86–94. https://doi.org/10.1016/j.yhbeh.2015.10.022 (2016).
34. Barrett, E. *et al.* Digit ratio, a proposed marker of the prenatal hormone environment, is not associated with prenatal sex steroids, anogenital distance, or gender-typed play behavior in preschool age children. *J. Dev. Orig. Health Dis.* https://doi.org/10.1017/S2040174420001270 (2020).
35. Golombok, S. & Rust, J. The Pre-School Activities Inventory: A standardized assessment of gender role in children. *Psycgol. Assess.* **5**, 131–136. https://doi.org/10.1037/1040-3590.5.2.131 (1993).
36. Bussey, K. & Bandura, A. Social cognitive theory of gender development and differentiation. *Psychol. Rev.* **106**, 676–713. https://doi.org/10.1037/0033-295X.106.4.676 (1999).
37. Endendijk, J. J., Groeneveld, M. G. & Mesman, J. The gendered family process Model: An integrative framework of gender in the family. *Arch. Sex. Behav.* **47**, 877–904. https://doi.org/10.1007/s10508-018-1185-8 (2018).
38. Morawska, A. The Effects of Gendered Parenting on Child Development Outcomes: A Systematic Review. *Clin. Child Family Psychol. Rev.* **23**, 553–576. https://doi.org/10.1007/s10567-020-00321-5 (2020).
39. Golombok, S. & Hines, M. Blackwell handbook of childhood social development. In *Blackwell handbooks of developmental psychology* (ed. Golombok, S.) (Blackwell Publishing, 2002).
40. Bandura, A. *Social Learning Theory* (General Learning Press, 1971).
41. Roychan, A., Verma, M., Podder, L. & Robert, S. Child Behavior: Role and Influence of Siblings. *Am. J. Nurs.* **12**, 13–18. https://doi.org/10.12691/ajnr-12-1-2 (2024).
42. Stoneman, Z., Brody, G. H. & MacKinnon, C. E. Same-sex and cross-sex siblings: Activity choices, roles, behavior, and gender stereotypes. *Sex Roles* **15**, 495–511. https://doi.org/10.1007/BF00288227 (1986).
43. Rust, J., Golombok, S., Hines, M., Johnston, K. & Golding, J. The role of brothers and sisters in the gender development of preschool children. *J. Exp. Child Psychol.* **77**, 292–303. https://doi.org/10.1006/jecp.2000.2596 (2000).
44. Hughes, C., McHarg, G. & White, N. Sibling influences on prosocial behavior. *Curr. Opin. Psychol.* **20**, 96–101. https://doi.org/10.1016/j.copsyc.2017.08.015 (2018).
45. Sang, S. A. & Nelson, J. A. The effect of siblings on children's social skills and perspective taking. *Infant Child Dev.* https://doi.org/10.1002/icd.2023 (2017).
46. Kaaz, T. & Heil, M. Infants' looking times in a 2-D mental rotation task. *Infant Child. Dev.* https://doi.org/10.1002/icd.2167 (2020).
47. Ernsten, L., Körner, L. M., Heil, M., Richards, G. & Schaal, N. K. Investigating the reliability and sex differences of digit lengths, ratios, and hand measures in infants. *Sci. Rep.* **11**, 10998. https://doi.org/10.1038/s41598-021-89590-w (2021).
48. Autometric Software Measurement of 2D: 4D Ratios (2004).
49. Ribeiro, E., Neave, N., Morais, R. N. & Manning, J. T. Direct Versus Indirect Measurement of Digit Ratio (2D:4D): A Critical Review of the Literature and New Data. *Evol. Psychol.* **14**, 1474704916632536. https://doi.org/10.1177/1474704916632536 (2016).
50. Leiner, D. J. *SoSci Survey (Version 3.1.06)* [Computer software], (2019). Available at https://www.soscisurvey.de.
51. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* 2nd edn. (Lawrence Erlbaum Associates, 1988).
52. Knickmeyer, R. C., Woolson, S., Hamer, R. M., Konneker, T. & Gilmore, J. H. 2D:4D ratios in the first 2 years of life: Stability and relation to testosterone exposure and sensitivity. *Horm. Behav.* **60**, 256–263. https://doi.org/10.1016/j.yhbeh.2011.05.009 (2011).
53. McIntyre, M. H., Ellison, P. T., Lieberman, D. E., Demerath, E. & Towne, B. The development of sex differences in digital formula from infancy in the Fels Longitudinal Study. *Proc. R. Soc. Lond. B Biol. Sci.* **272**, 1473–1479. https://doi.org/10.1098/rspb.2005.3100 (2005).
54. Butovskaya, M. *et al.* Sex, population origin, age and average digit length as predictors of digit ratio in three large world populations. *Sci. Rep.* **11**, 8157. https://doi.org/10.1038/s41598-021-87394-6 (2021).
55. Ernsten, L. *et al.* The association of prenatal amniotic sex hormones and digit ratio (2D:4D) in children aged 5 to 70 months: A longitudinal study. *PLoS One* https://doi.org/10.1371/journal.pone.0282253 (2023).
56. Trivers, R., Manning, J. & Jacobson, A. A longitudinal study of digit ratio (2D:4D) and other finger ratios in Jamaican children. *Horm. Behav.* **49**, 150–156. https://doi.org/10.1016/j.yhbeh.2005.05.023 (2006).
57. Lolli, L. *et al.* A comprehensive allometric analysis of 2nd digit length to 4th digit length in humans. *Proc. R. Soc. B Biol. Sci.* **284**, 20170356. https://doi.org/10.1098/rspb.2017.0356 (2017).
58. Forstmeier, W. Avoiding misinterpretation of regression lines in allometry: Is sexual dimorphism in digit ratio spurious?. *Biorxiv* https://doi.org/10.1101/298786 (2018).
59. Manning, J. T. *Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease* (Springer, 2012).
60. Manning, J. T. & Fink, B. Sexual dimorphism in the ontogeny of second (2D) and fourth (4D) digit lengths, and digit ratio (2D:4D). *Am. J. Hum. Biol.* https://doi.org/10.1002/ajhb.23138 (2018).
61. Kuzawa, C. W. Adipose tissue in human infancy and childhood: An evolutionary perspective. *Am. J. Phys. Anthropol.* **107**, 177–209. https://doi.org/10.1002/(SICI)1096-8644(1998)107:27+%3c177::AID-AJPA7%3e3.0.CO;2-B (1998).
62. McIntyre, M. H., Cohn, B. A. & Ellison, P. T. Sex dimorphism in digital formulae of children. *Am. J. Phys. Anthropol.* **129**, 143–150. https://doi.org/10.1002/ajpa.20240 (2006).
63. Braun, S. S. & Davidson, A. J. Gender (Non) conformity in middle childhood: A mixed methods approach to understanding gender-typed behavior, friendship, and peer preference. *Sex Roles* **77**, 16–29. https://doi.org/10.1007/s11199-016-0693-z (2017).
64. Langlois, J. H. & Downs, A. C. Mothers, Fathers, and Peers as Socialization Agents of Sex-Typed Play Behaviors in Young Children. *Child Dev.* **51**, 1237–1247. https://doi.org/10.2307/1129566 (1980).

65. Fagot, B. I. The influence of sex of child on parental reactions to toddler children. *Child Dev.* **29**, 459–465 (1978).
66. Berenbaum, S. A. & Beltz, A. M. How early hormones shape gender development. *Curr. Opin. Behav. Sci.* **7**, 53–60. https://doi.org/10.1016/j.cobeha.2015.11.011 (2016).
67. Berenbaum, S. A., Blakemore, J. E. O. & Beltz, A. M. A Role for Biology in Gender-Related Behavior. *Sex Roles* **64**, 804–825. https://doi.org/10.1007/s11199-011-9990-8 (2011).
68. Manning, J. T., Churchill, A. J. G. & Peters, M. The effects of sex, ethnicity, and sexual orientation on self-measured digit ratio (2D:4D). *Arch.Sex. Behav.* **36**, 223–233. https://doi.org/10.1007/s10508-007-9171-6 (2007).
69. Beltz, A. M., Kelly, D. P. & Berenbaum, S. A. *Neural Circuit and Cognitive Development* (Academic Press, 2020).
70. Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R. & Manning, J. T. 2nd. to 4th digit ratios, fetal testosterone and estradiol. *Early Hum. Dev.* **77**, 23–28. https://doi.org/10.1016/j.earlhumdev.2003.12.002 (2004).
71. Richards, G., Browne, W. & Constantinescu, M. Digit ratio (2D:4D) and amniotic testosterone and estradiol: An attempted replication of Lutchmaya et al. (2004). *J. Dev. Orig. Health Dis.* https://doi.org/10.1101/2020.07.10.197269 (2020).
72. Ventura, T., Gomes, M. C., Pita, A., Neto, M. T. & Taylor, A. Digit ratio (2D:4D) in newborns: influences of prenatal testosterone and maternal environment. *Early Hum. Dev.* **89**, 107–112. https://doi.org/10.1016/j.earlhumdev.2012.08.009 (2013).
73. Sorokowski, P. & Kowal, M. Relationship between the 2D:4D and prenatal testosterone, adult level testosterone, and testosterone change: Meta-analysis of 54 studies. *Am. J. Biol. Anthropol.* **183**, 20–38. https://doi.org/10.1002/ajpa.24852 (2024).

Author contributions

L.E., L.M.K., M.H., and N.K.S. contributed to the study idea and design. L.E. performed the primary analysis, and L.M.K., M.H., and N.K.S. verified the analytical methods. L.E. designed the tables and drafted a first version of the manuscript. L.M.K., M.H., and N.K.S. provided critique to refine the manuscript. All authors approved the final version of this manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Study 4:

Ernsten, L., Seehagen, S., Nonnenmacher, N., Wolf, O. T., Heil, M., Schaal, N. K. (2024). *Preliminary findings on the association between maternal salivary and hair cortisol and the mother-infant-interaction during the early postpartum period.* Manuscript accepted for publication after minor revision in Psychoneuroendocrinology.

Author contributions: Ernsten, L.: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – Original Draft, Visualization; Seehagen, S.: Writing – Review & Editing; Nonnenmacher, N.: Writing – Review & Editing; Tautz, J.: Writing – Review & Editing; Wolf, O. T.: Writing – Review & Editing; Heil, M.: Methodology, Writing – Review & Editing; Schaal, N. K.: Conceptualization, Methodology, Writing – Review & Editing, Supervision

Impact Factor (2022): 3.7 (Journal Citation Reports, 2023b)

Preliminary findings on the association between maternal salivary and hair cortisol and the mother-infant-interaction during the early postpartum period

Luisa Ernsten [a,*], Sabine Seehagen [b], Nora Nonnenmacher [c], Juliane Tautz [d], Oliver T. Wolf [b], Martin Heil [a], Nora K. Schaal [a]

[a] Institute for Experimental Psychology, Heinrich-Heine-University Düsseldorf, Germany

[b] Faculty of Psychology, Ruhr-University-Bochum, Germany

[c] Institute for Medical Psychology, Center for Psychosocial Medicine, Heidelberg University Hospital, Germany

[d] Clinic for General Pediatrics, Neonatology and Pediatric Cardiology, University Clinic Düsseldorf, Germany

* Corresponding author:
Luisa Ernsten
Institute of Experimental Psychology
Heinrich-Heine University Düsseldorf
Universitätsstraße 1
40225 Düsseldorf
Germany
+49 (0) 211/ 81-14566

Abstract

Maternal capabilities to engage in sensitive caregiving are important for infant development and mother-infant-interaction, however, can be negatively affected by cortisol due to a stress response. Previous research suggested that cortisol possibly impairs cognitive functions important for caregiving behavior, which potentially leads to less maternal sensitivity. However, studies investigating the influence of cortisol using endocrine parameters on the mother-infant-interaction during the early postpartum are lacking. In the current study, fifty-nine mother-infant-dyads participated in a laboratory face-to-face still-face (FFSF) observation when infants were 4 months of age. Maternal and infant positive, negative and matched behavior during the FFSF was microanalytically coded. Cortisol concentrations were obtained using hair and saliva samples. For salivary cortisol, the area under the curve with respect to ground (AUC_G) was calculated using two saliva samples obtained after arrival and after the FFSF. Multiple block-wise hierarchical linear regression models were conducted to incorporate potential confounding factors (maternal age, parity, infant gestational age, infant sex) in a first step and, then, test for the association of hair and salivary cortisol with maternal and infant positive, negative and dyadic behavior in a second step. For both it was hypothesized that cortisol assessed in hair and saliva is negatively associated with positive and matched mother-infant-interaction, and positively associated with negative mother-infant-interaction. It could be shown that salivary but not hair cortisol as well as infant gestational age and infant sex related significantly to infant positive and negative affect as well as matched behavior during the reunion phase of the FFSF. Maternal positive affect was unrelated to any of the variables. The results are discussed in regard to the importance of maternal cortisol levels over a longer period of time and more acute situational levels for the mother-infant-interaction as well as the relevance of included confounding factors.

Keywords: HPA axis, stress, salivary cortisol, hair cortisol, still-face paradigm, mother-infant-interaction

Study registration: DRKS00024921 (registered on April 27th, 2021)

Word count: 8,166

1. Introduction

The mother-infant-interaction is – besides the interaction with other relevant caregivers – one of the first experiences of newborn infants and crucial for infant's brain development (Ilyka et al., 2021). The mother serves in a coregulating function as infants cannot yet fully cope with different stimuli (Laurent et al., 2016), and enables the development of self-regulation strategies (Blair & Ku, 2022). For this, however, the mother needs to adequately perceive, interpret and react to the infant's signals, a skill called maternal sensitivity (Bowlby & Ainsworth, 2000; Shin et al., 2008). Women transitioning to motherhood face crucial changes which could potentially be stressful for the mother (Emmanuel & St John, 2010). Although, evidence suggests a negative effect of psychosocial stress on maternal sensitivity during mother-infant-interactions (Almanza-Sepulveda et al., 2020; Booth et al., 2018; Finegood et al., 2016), the effects of cortisol, a hormone that is released by the hypothalamus pituitary adrenal (HPA) axis and can also be associated with psychosocial stress (Smith & Vale, 2006; Tsigos et al., 2020), on maternal abilities to engage in sensitive caregiving have not been extensively examined.

The HPA axis regulates different bodily functions and underlies a diurnal rhythm where cortisol levels rise right after waking up and then progressively decrease during the course of the day (Bhake et al., 2019; Ranjit et al., 2005). In the context of an occurring stressor, the HPA axis is activated (further referred to as the "stress response") and the hypothalamus releases corticotropin releasing hormone, which induces the release of adrenocorticotrophic hormone from the pituitary gland, ultimately causing the secretion of cortisol by the adrenal gland into the body, which prepares the individuum for a subsequent stress response (e.g., fight or flight; Cole & Kramer, 2016). The secretion of cortisol following a stress response is typically terminated due to a negative feedback loop that regulates the synthesis of cortisol (Smith & Vale, 2006). If a stressor persists a dysregulation or even alteration of the HPA axis activity and the diurnal cortisol slope have been reported as a possible consequence (McEwen, 2004; Sjögren et al., 2006; Tsigos et al., 2020).

It has been discussed that cortisol affects different cognitive systems, specifically, when elevated following a stressor. Plessow et al. (2011) and Goldfarb et al. (2017) discussed the effect of cortisol elevation following a stressor on cognitive flexibility. Plessow et al. (2011) found that cortisol led to a decrease in cognitive flexibility and stressed individuals had more difficulties compared to non-stressed individuals to include new information in already existing information, while Goldfarb et al. (2017) found that stressed individuals had more difficulties to shield themselves from distractors. Whereas Plessow et al. (2011) used a psychosocial stress task, the Trier Social Stress Test (Kirschbaum et al., 2008), Goldfarb et al. (2017) used a cold pressure task. Although both stress task can evoke a cortisol elevation, it has been discussed that the elevation of HPA axis activity due to a stress response seems to be more sensitive towards psychosocial stressors (Pruessner & Ali, 2015). However, both studies found that stressed individuals did not show worse overall performance compared to non-stressed individuals (Goldfarb et al., 2017; Plessow et al., 2011). Schwabe and Wolf (2013) discussed that cortisol elevation possibly affects memory systems that ultimately lead to a shift from cognitive (i.e., sensitive) to more habitual memory systems. As the authors note, this may lead to less goal-directed behavior, which can affect behavioral flexibility, however, can also be associated with a less demanding memory function which allows the individuum to use the free resources to cope with an occurring stressor (Schwabe & Wolf, 2013). Mother's ability to sensitively interact with an infant potentially relies on a manifold of factors. As defined by Shin et al. (2008), maternal sensitivity relies on maternal perception and interpretation of infant cues as well as on an appropriate response in a contingent manner. Further, sensitive mothers should be able to adapt and change these processes during the mother-infant-interaction, which makes maternal sensitivity a highly dynamic construct (Shin et al., 2008). Based on this definition, it is likely that cortisol can affect cognitive processes that are relevant for maternal sensitivity. Further, if an individuum cannot cope with a stressor over a longer period leading to a dysregulation of the HPA axis, an alternated responsiveness of the HPA axis can be the consequence which can be associated with various risks for mental health problems and psychiatric disorders (Hammen, 2005; McEwen, 2004).

Based on the aforementioned literature (Goldfarb et al., 2017; Plessow et al., 2011; Schwabe & Wolf, 2013) cortisol can possibly also impair maternal capabilities to engage in sensitive caregiving, specifically, at the cost of cognitive and behavioral flexibility. However, flexibility could be crucial for mothers looking at some important demands during dyadic interaction. The mother-infant-interaction can be characterized by states of matches and mismatches between both interaction partners (DiCorcia & Tronick, 2011). During matching states, maternal regulatory input aligns with infant regulatory needs, while failing to do so can lead to mismatching states (DiCorcia & Tronick, 2011). At early infancy, maternal sensitivity helps the mother to adequately repair the interaction and change back to matching states when infants are not capable to regulate themselves (DiCorcia & Tronick, 2011; Gianino & Tronick, 1988). The effective dyadic reparation after a mismatching state fosters infant's own coping capabilities and promotes the development of self-regulatory strategies, whereas ineffective reparation potentially stresses the infant which is assumed to hinder infants' socio-emotional and cognitive development (Braungart-Rieker et al., 2014; DiCorcia & Tronick, 2011). Maternal sensitivity could predict infant's responses to stressors and dyadic regulation in standardized paradigms like the face-to-face still-face (FFSF; Mesman et al., 2009). During the FFSF, caregivers provoke a mismatching state making it possible to examine the infant's reaction and dyadic patterns of regulation (Mesman et al., 2009; Tronick et al., 1978). As a reaction to the caregiver's still-face, the infant reliably increases negative affect, protesting and self-comforting behavior while decreasing positive affect (the so-called still-face effect), and the dyadic regulation can be reflected in a change back to affective states comparable to the play phase after the mother-infant-interaction is resumed (Mesman et al., 2009). It could be shown that infant's responses and dyadic regulation can be associated with the quality of attachment (Müller et al., 2022), maternal sensitivity (Braungart-Rieker et al., 2014), and psychosocial stress (Tronick et al., 2021). Likewise, elevated cortisol levels over a longer period of time may be associated with maternal mental health problems, like depression or anxiety, which have been shown to affect the quality of mother-infant-interactions (Feldman et al., 2009) and maternal sensitivity (Stanley et al., 2004; Tester-Jones et al., 2016).

There are only few studies that examined the association between maternal cortisol and maternal sensitivity during mother-infant-interactions in the early postpartum period. During pregnancy until 3 months postpartum, maternal cortisol levels are generally higher compared to non-pregnant women, and change back to normal levels after 3 months postpartum (Almanza-Sepulveda et al., 2020). In humans, higher cortisol levels shortly after birth could be associated with higher alertness for infant cues, however, after 3 months postpartum, higher levels of cortisol could be associated with less optimal caregiving and sensitivity in mothers (Almanza-Sepulveda et al., 2020), with the latter possibly adding to the existing literature on the effects of elevated cortisol in the context of a stress response. Studies considering salivary cortisol concentrations (SCC) as a marker for situational cortisol levels (Kirschbaum & Hellhammer, 2008) during mother-infant-observations are scarce. Thompson and Trevathan (2008) found that maternal cortisol elevation during a free-play observation led to less synchronous behavior between mothers and their 3-month-old infants, which was interpreted as less maternal sensitivity by the authors. Finegood et al. (2016) found that mothers who experienced adverse life events showed a negative association between cortisol levels and maternal sensitivity during the first 2 years after birth during free-play observations at 7 and 15 months of age and a puzzle-task at 24 months of age. However, these studies slightly differ regarding the assessment of salivary cortisol, with Thompson and Trevathan (2008) using two measures during the observation and calculating the difference between the obtained cortisol levels to index a cortisol increase or decrease, whereas Finegood et al. (2016) obtained three samples over the course of the experiment and calculated the average of all samples. Further, none of these studies used the FFSF during observations, although it allows a more standardized observational method to assess maternal sensitivity. Another study indicated that a maternal stressor prior to the FFSF potentially leads to more negative interaction patterns between the mother and the infant, however, did not control for maternal cortisol levels following the stress manipulation (Mueller et al., 2021; Tronick et al., 2021). Using human hair for the examination of cumulated cortisol levels (Meyer & Novak, 2012), Khoury et al. (2020) found higher hair cortisol concentrations (HCC) during pregnancy and depressive symptoms

of the mother related to maternal withdrawal and intrusiveness during the FFSF with 4-months-olds. Tarullo et al. (2017) found a negative association between maternal postpartum HCC and positive engagement, and a positive association with maternal intrusiveness during the FFSF with 5 to 7-months-olds. In sum, both long-term as well as situational cortisol levels seem to be important factors that can affect the mother-infant-interaction. To date no prior study investigated both markers, HCC as a marker for cumulated cortisol levels and SCC as a marker for situational cortisol levels, and the mother-infant-interaction during the FFSF.

In sum, the mother-infant-interaction is important for the socio-emotional development of infants and is affected by maternal abilities to engage in sensitive caregiving. Although, maternal cortisol levels possibly affect these maternal abilities and related cognitive systems (Almanza-Sepulveda et al., 2020; Booth et al., 2018; Goldfarb et al., 2017; Plessow et al., 2011; Schwabe & Wolf, 2013), few studies investigated the influence of situational and long-term cortisol levels, in the early postpartum period. Endocrine markers can be used to assess situational and long-term cortisol levels as indicators for overall HPA axis activity as well as for a stress response (Hellhammer et al., 2009; Kirschbaum & Hellhammer, 2008; Stalder & Kirschbaum, 2012; Stalder et al., 2017) and significant associations between SCC and HCC with maternal sensitivity or the quality of mother-infant-interactions have been found (Finegood et al., 2016; Khoury et al., 2020; Nystrom-Hansen et al., 2019; Tarullo et al., 2017; Thompson & Trevathan, 2008). The current study aims to investigate the association between maternal situational and cumulated cortisol levels with the mother-infant-interaction during the FFSF in a sample of mother-infant-dyads 4 months postpartum. The FFSF provokes a mismatching state between mothers and their infants. During the reunion, mothers resume the normal interaction and dyadic regulation can be observed as a consequence of the mismatching state. The first part of our study analyzed the so-called still-face effect with the assumption that infants react to the mismatching state with less positive affect, more negative affect and more self-comforting behavior while mothers maintain a still-face. During the last phase of the FFSF, it is proposed that infants negative and positive affect as well as self-comforting behavior

resumes to levels comparable to the start of the FFSF. To reflect maternal cortisol, two measures were used. SCC was assessed and should reflect more momentary maternal cortisol levels during the interaction with her infant. HCC was used to reflect long-term, cumulated maternal cortisol levels during the past 3 months. Both measures are therefore related to activity of the HPA axis and evidence suggests moderate to large correlations between SCC and HCC (Short et al., 2016; Singh Solorzano et al., 2023; Zhang et al., 2018). The mother-infant-interaction is operationalized by different maternal, infant and dyadic behaviors assessed with the Infant and Caregiver Engagement Phases Revised (ICEP-R; Reck et al., 2009). The ICEP-R allows to individually assess maternal affect, infant affect and the dyadic behavior of mothers and infants (Reck et al., 2009). In the current study, higher SCC and HCC are both hypothesized to be associated with less positive infant and maternal affect, more negative infant affect, less matched dyadic behavior, which was operationalized as shared positive affect and eye-contact, as well as with a longer delay of dyadic regulation after maternal unresponsiveness during the FFSF indicated by shared positive affect.

2. Methods

The current study is part of a larger monocentric, prospective cohort study. Data collection occurred between August 2021 and January 2022 at the Institute of Experimental Psychology at the Heinrich-Heine-University Düsseldorf. The study protocol was approved by the local ethics committee (No. 2021-1329) and in accordance with the declaration of Helsinki. The overall study aims to investigate the differential effects and interactions between maternal and infant cortisol and maternal perceived stress as well as maternal mental health on the mother-infant-interaction as well as bonding with a specific focus on differences between mothers with full-term and preterm infants. As recruitment of the preterm mothers was far more difficult than anticipated, data collection of the overall study is currently still ongoing. The current work is a secondary analysis of the whole data set that only includes full-term infants

and their mothers. Further, the current work focuses only on maternal cortisol measures and the association with the mother-infant-interaction.

2.1. Recruitment and participants

Mother-infant-dyads were recruited via the residents' registration office of the city of Düsseldorf, Germany. Members of the study group contacted mothers who indicated interest in study participation via telephone and informed them about the purpose and procedure of the study. All mothers had to be fluent in German and over 18 years old. Infants had no serious illness or congenital developmental disorders. Dyads were invited when infants were 4 months of age (see Table 1 for mean age). All legal guardians of infants gave their written informed consent for study participation, data recording and storage and received a travel refund as well as a present for participation at the end of the experiment.

In total, *N* = 76 mothers with full-term infants participated in the study. Of these dyads, *n* = 8 had to be excluded due to technical issues or a deviation from the study protocol, *n* = 5 had to be excluded as infants expressed crying or protesting behavior ≥80% of the overall time during the FFSF, and of *n* = 1 dyad the corresponding SCC could not be analyzed, leaving a sample of *n* = 62 dyads. For *n* = 3 dyads, the corresponding HCC was regarded as a statistical outlier, as well as for *n* = 1 of these the corresponding SCC, and were not included in the corresponding analyses, leaving a final sample of *n* = 59 for data analysis. All sociodemographic variables of included dyads are reported in Table 1.

Table 1

Overview of maternal and infant demographic variables.

	N	*M*	*SD*	*Min*	*Max*
Mothers					
Age (in years)		35.41	4.32	26	43
Nationality					
German	52				
Other	7				

Marital status					
Married/registered partnership and living together	44				
Married/registered partnership and not living together	1				
Living with a partner	9				
Single	5				
Education					
Technical diploma	1				
High-school diploma	10				
University degree	41				
Doctorate	7				
Parity					
Primiparous	31				
Multiparous	28				
Infants					
Sex assigned at birth					
Male	31				
Female	28				
Gestational age at birth (in days)		277.58	8.64	262	292
Age at study participation (in days)		127.12	6.52	111	143
Birth weight (in g) [a]		3492.67	388.95	2640	4460
Birth height (in cm) [a]		52.06	2.06	48	57
Birth mode					
Vaginal	41				
C-section	12				
Assisted	6				

Note: N = 59 mother-infant-dyads; [a] data of n = 58 due to missing information of one dyad.

2.2. Materials and measures

2.2.1. Maternal hair cortisol concentration

To assess maternal cumulative cortisol levels via HCC, a standardized protocol for hair collection and storage was used (Meyer et al., 2014). At the end of the experiment, the hair was divided from a posterior vertex position and, using a loop, an approximately 3 mm thick hair strand was tied off. The strand was ultimately cut off closely to the scalp and wrapped in tin foil. All samples were separately stored in air-tight bags. Until analysis, all samples were stored dark and at room temperature. Analyses were carried out by the laboratory of the Department of Cognitive Psychology at the University of Bochum. Hair strands were washed,

dried, and grinded prior to cortisol extraction. The exact method can be seen elsewhere (Meyer et al., 2014). For analysis, the commercially available chemiluminescence immunoassay with high sensitivity (IBL Internation GmbH, catalogue number R62111) was used. The intra and inter assay coefficients of variance are reported to be 0.4-1.7% and 0.8-1.8% respectively (IBL Internation GmbH, catalogue number R62111). Human hair has an average growth rate of 1 cm per month and the cortisol concentration obtained refers analogously to one month (Meyer et al., 2014). In the current study, the first 3 cm were analyzed and refer to the cumulated HCC during the past 3 months in pg/mg. Mothers had a mean HCC of *M* = 4.19 pg/mg (*SD* = 2.20 pg/mg, range 1.40-11.36 pg/mg).

2.2.2. Salivary cortisol concentrations

Salivary cortisol concentrations were measured for both mothers and infants and were assessed at three time points during the assessment: right at the beginning before behavioral observations (T0), right after behavioral observations (T1), and approximately 25 minutes after T1 (T2). For the current study, only maternal SCC of T0 and T1 were of interest. Assuming a latency of approximately 20 minutes (Kirschbaum & Hellhammer, 2008), T0 should reflect the baseline cortisol levels of mothers before the examination, T1 approximately at arrival, and T2 right after the mother-infant-observation. Therefore, T0 and T1 should reflect maternal cortisol levels at the beginning of the examination, and should be unaffected by any circumstances that occurred during the FFSF (e.g., infant crying). For analysis, the area under the curve with respect to ground (AUC_G) was calculated based on the recommendations of Pruessner et al. (2003). The AUC_G reflects a time-adjusted mean of two measurements which should reflect maternal baseline cortisol levels prior and during the FFSF. The interval in minutes between T0 and T1 was *M* = 20.51 (*SD* = 9.11), and between T1 and T2 *M* = 32.60 (*SD* = 9.54). Mothers were instructed to thoroughly insalivate a cotton swab. For the infants, mothers held the cotton swabs in their infants' mouth to collect saliva until the cotton swab was visibly insalivated. Saliva samples were frozen and stored at -20 degrees Celsius until analysis. Analyses were carried out at Dresden LabService GmbH, Technical University of Dresden, Germany. After

thawing, samples were centrifuged at 3,000 rpm for 5 min, which resulted in a clear supernatant of low viscosity. SCC were measured using commercially available chemiluminescence immunoassay with high sensitivity (IBL Internation GmbH, catalogue number R62111). The intra and inter assay coefficients of variance were below 9%. Cortisol concentrations are reported in nmol/l. Mothers had a mean SCC at T0 of *M* = 3.52 nmol/l (*SD* = 2.56 nmol/l, range 0.73-14.47 nmol/l), T1 of *M* = 3.31 nmol/l (*SD* = 2.28 nmol/l, range 0.63-10.89 nmol/l), and T2 of *M* = 2.51 nmol/l (*SD* = 1.60 nmol/l, range 0.59-9.48 nmol/l). As free cortisol substantially fluctuates during the course of the day (Bhake et al., 2019; Ranjit et al., 2005), time of day is an important factor while investigating SCC. Given a diurnal slope with highest cortisol levels shortly after waking up and a decline approximately at midday, time of day of sampling will be incorporated in the current study. The analyses of interested were tested again incorporating time of day as a dummy coded variable with '0 = before 12 pm' and '1 = after 12 pm'. As time of day as an additional factor did not change the initial findings, results can be seen in the supplementary material (see Tables S5-S11).

2.2.3. Coding of maternal affect, infant affect and dyadic behavior

The mother-infant-observations were analyzed using Mangold INTERACT software and the behavior was coded according to the German version of the Infant and Caregiver Engagement Phases Revised (ICEP-R; Reck et al., 2009) by two trained raters. It allows to microanalytically and separately code infant and maternal behavior as well as dyadic behavior during the interaction. Infant behavior can be coded as negative engagement (i.e., withdrawn and protest), object/environment engagement, social monitor, and social positive engagement. Maternal behavior can be coded as negative engagement (i.e., withdrawn, hostile and intrusive), non-infant focused engagement, social monitor/no vocalizations or neutral vocalizations, social monitor/positive vocalizations, and social positive engagement. Dyadic behavior was coded as dyadic eye-contact and joint activity/looking. Behavior was coded in seconds, indicating the total amount of time during the phases the respective code was shown. Infant, maternal and dyadic codes were mutually exclusive. Furthermore, additional

information about self-comforting behavior (oral and self-clasp), distancing, and autonomic stress indicators of the infant as well as rough touches or violations of the FFSF of the mother were coded. For each code (except for *interactive reparation* which is reported in seconds), the total duration during a phase was then divided by the total length of the phase indicating the relative amount of the exhibited behavior during the respective phase. Raters were trained to reach 80% reliability prior to coding. Ten videos were randomly selected to test for inter-rater-reliability and were coded by an additional third coder. For infant codes, the time-unit kappa was .76-.78 (agreement 86-88%), and event-alignment kappa was .51 (agreement 79%). For maternal codes, time-unit kappa was .50-.54 (agreement 63-66%), and event-alignment kappa was .39 (agreement 66%). Dyadic codes reached a time-unit kappa of .69-.70 (agreement 92%), and event-alignment kappa of .56 (agreement 81%). Table 2 gives a brief description of the behavior of interest for the current study. In any cases where raters had difficulties or were uncertain about which code to assign in mother-infant-observations, the team of raters and the coding trainer discussed these instances to determine the appropriate code for the respective behavior observed in the video. For all codes, the relative amount of exhibited behavior was used for analysis, that is the total duration of the behavior during a phase was divided by the total length of the respective phase. See supplemental Table S 1 for the duration of used codes in means and standard deviations.

Table 2

Description of codes used for maternal, infant and dyadic behavior.

Code	Description
Infant social positive engagement [a]	Infants look at their caregiver and exhibit overall positive vocalizations, expressions and behavior. These include smiling, cooing, and laughing.
Infant social monitor [a]	Infants look at the caregiver, however, do not exhibit explicit positive behaviors. Infants have a neutral or interested facial expression.
Infant protesting behavior [a]	Infants are protesting and exhibit overall negative vocalizations, expressions and behavior. These include crying or fussing, as well as overall signs of anger and frustration. Infants are overall active, expressed by arching the back, kicking arms against the chair, trying to escape or to get picked up, as well as pushing and pulling away from caregivers.
Infant self-comforting behavior [a]	*Oral:* Infant stimulate themselves by sucking on their body, objects, or the caregiver's hand or fingers. There needs to be skin contact between infant's mouths and the body, objects, or

	caregivers. For sucking on their own body or objects, the behavior needs to be initiated by infants, for sucking on caregiver's hands or fingers, there is no rule for coding in terms of initiation. *Self-clasp:* There is contact between the hands and/or fingers of infants.
Infant positive affect [b]	This code is the sum of *infant social positive engagement* and *infant social monitor*.
Maternal positive affect [b]	This code is the sum of *maternal social positive engagement* and *social monitor/ positive vocalizations* (cf. Müller et al., 2022). Maternal *social positive engagement* is characterized by overall positive expressions, vocalizations and behavior. The mother laughs or smiles, expresses play faces, and talks to the infant in a positive manner or uses infant-directed speech. Maternal *social monitor/ positive vocalizations* is characterized by more neutral expressions and behavior, however, vocalizations are positive, i.e., mothers use infant-directed speech or make positive sounds like kissing or clicking. Her facial expression is neutral or interested. The mother is overall focused on her infant's activities.
Match [b]	The total amount of time mothers and infants display *social positive engagement* and/or *social monitor* (for mothers: *social monitor/ positive vocalizations*, see descriptions above) at the same time.
Interactive reparation [b]	The latency from the beginning of the reunion phase to the first onset of a matching state (*match*). This code indicates how long it takes for the dyad to restore the interaction and changing into a matching state after the disruption due to the maternal still-face. This code is given in seconds.
Dyadic eye-contact [b]	The mother and infant are looking at each other's face.

Note: All codes refer to the description in Infant and Caregiver Engagement Phases Revised (ICEP-R) Heidelberg Version by Reck et al. (2009). [a] These codes were used to test changes in infant affect and behavior during maternal unresponsiveness and after re-engagement. [b] These codes were used to test the association between SCC and HCC with maternal affect, infant affect and dyadic behavior during the last episode of the FFSF.

2.2.4. Maternal depressive symptoms

Maternal depressive symptoms were assessed using the German version of the Edinburgh Postnatal Depression Scale (EPDS; Bergant et al., 1998) as part of the questionnaire mothers filled out after the video observation. The EPDS was originally developed by Cox et al. (1987) to assesses maternal depressive symptoms on ten items that can be answered on a 4-point Likert-scale (0 to 3) and are summed up to a total score (range 0-30) with higher scores indicating more severe maternal depressive symptoms (Bergant et

al., 1998; Cox et al., 1987). Different cut-off scores have been discussed to screen mothers for clinically relevant symptoms of a major depression (Cox, 2019) and the German validation study of Bergant et al. (1998) suggested an EPDS score of >9.5 to screen for mild depressive symptoms and a cut-off >12.5 for the diagnosis of a major depressive disorder. The German version of the EPDS revealed good split-half reliability of .82 and good internal validity of Cronbach's α = .81 (Bergant et al., 1998). In the current study, EPDS scores were available for *n* = 57 mothers (two mothers did not answer item 8) and mothers had mean EPDS scores of *M* = 6.11 (*SD* = 3.89, range 0-18). The analyses of interest were again tested incorporating the EPDS as an additional factor. As the incorporation of the EPDS did not change the initial findings, the results can be seen in the supplementary material (see Tables S5-S11).

2.3. Procedure

Mother-infant-dyads came to the laboratory and were informed about the overall study procedure prior to obtaining the written informed consent. They gave the first saliva sample before starting the video observation. Mothers and infants were seated in front of each other and infants were fastened in a baby chair that was slightly elevated so that mothers and infants were approximately on eye level. Two cameras recorded maternal and infant behavior simultaneously. The examiner was in the room behind a black curtain throughout the whole procedure and preserved silent during the observation. Mothers and infants were shielded by two partition walls during the observation. Prior to the FFSF, mothers and infants engaged in a 5-minute free play to familiarize with the setting, where mothers were instructed to engage with their infants as normally and naturally as possible. They were allowed to touch their infant, however, not to lift them out of the seat. Mothers were given an acoustic signal (i.e., a ringing bell) to indicate the start and end of the 5-minute period. Afterwards, the examiner briefly repeated the explanation of the FFSF to make sure the mother knew how to act during the procedure. The standardized FFSF (Tronick et al., 1978) consisted of three 120-seconds-periods: a play, still-face, and a reunion phase. In the play phase, mothers should continue the normal interaction with their infant. Following an acoustic signal, the mothers were instructed

to turn their head to the side, to adopt a neutral (still-) face, and shift the head back towards the infant, however, looking at a point slightly above the infant's head. They were then instructed to cease all touching and to respond to none of the infants' signals by maintaining their neutral (still-) face for the following 120 seconds. After a further signal, mothers were allowed to resume the normal interaction with their infant. During the whole procedure, mothers were not allowed to lift their infant out of the seat nor to use toys. If the mother did not follow the protocol, the examiner gave a brief instruction how to act (e.g., to remain a neutral face or to engage with the infant). The examiner shortened the phase if the infant cried for 15 seconds non-stop. In the current study, play, still-face, and reunion phase had mean durations (with standard deviations in parentheses) in seconds of 117.45 (10.94), 119.48 (11.13), and 117.99 (9.57) respectively. After completion of the FFSF, mothers and infants salivated the second saliva sample. Then the mother was handed out the questionnaires. She gave information about her and the infants' social demographics (i.e., age, nationality, marital status, educational levels, infants' gestational age at birth, infant sex assigned at birth, birth weight and height), about her pregnancy (i.e., mode of delivery) and birth experience, and she completed standardized questionnaires, among others (that are not included int the current study) the EPDS. After approximately 25 minutes, the last saliva sample was gathered from both mother and infant. Lastly, the examiner cut off the hair strand of the mother. After completion, mothers received the travel refund and a child's rattle as a gift for the infant.

2.4. Statistical analysis

An a priori power analysis was conducted using G*Power (version number 3.1.9.6) to determine the minimum sample size required to achieve a medium effect size of $f^2 = .25$ in the larger study based on the study results of Tarullo et al. (2017), who investigated the association of HCC and maternal and infant behavior during a free play. To achieve 80% power and alpha-levels set to .05, the minimum sample size for a Pearson correlation was $N = 97$. For the current study, a post-hoc power-analysis was run to investigate the achieved power for the current subsample ($N = 59$). Looking at block-wise hierarchical multiple linear regression

models (fixed model, R^2 increase) that reached statistical significance and alpha-levels set to .05, the achieved power in the current study ranged between .79 and .84 (depending on the criterion).

All analyses were performed using the statistical software SPSS (version number 28.0.1.0). All alpha-levels were set to .05. We tested for the assumed still-face effect of infants during the FFSF to check for infant's reaction towards maternal unresponsiveness, using repeated measures ANOVA with the within-subjects factor *phase* (play vs. still-face vs. reunion) and post-hoc Bonferroni-adjusted pair-wise comparisons on the relative duration of *infant social positive engagement*, *infant protesting behavior* and *infant self-comforting behavior* (*oral* and *self-clasp*) during the respective phase. A decrease of *infant social positive engagement* as well as an increase of *infant protesting behavior* and *infant self-comforting* (*oral* and *hand-grasping*) behavior from play to still-face, and an increase of *infant social positive engagement* and a decrease of *infant protesting behavior* from still-face to reunion phase was assumed.

To test our hypotheses that maternal SCC and HCC relate to less *infant positive affect*, higher *infant protesting behavior,* less *maternal positive affect,* less *match*, less *interactive reparation*, and less *dyadic eye-contact*, different block-wise hierarchical multiple linear regression models were performed. Prior to the regression models, Spearman correlations were conducted to test for the association between the study variables. For all models, in a first step the possible confounding factors *maternal age*, *parity* (dummy-coded as '0 = primiparous', '1 = multiparous'), *infant gestational age* and *sex assigned at birth* (dummy coded as '0 = female', '1 = male'), were included to control for possible influences. As the sample was relatively homogeneous concerning (relatively high) educational levels (approximately 83% of women with a university degree or higher, see Table 1), this variable was not included. Next, the additional variables HCC and SCC were added in a second model. As HCC and SCC were both not normally distributed, log-transformations with base 10 were performed prior to data analysis for HCC and the AUC_G of SCC. Additional analyses were conducted to test the same

hierarchical regression models as noted, however, incorporating time of day of saliva collection and maternal depressive symptoms obtained with the EPDS as additional factors to control for in the second model. The results of these analyses can be found in the supplementary material as the initial results were not changed due to the addition of these two factors (see Table S5-11).

All effects were interpreted according to Cohen (1988). For η^2_p this means η^2_p = .01 is interpreted as a small effect, η^2_p = .06 as a medium effect, and η^2_p = .14 as a large effect. Correlations were interpreted as r = .10 as a small effect, r = .50 as a medium effect, and r = .50 as a large effect. Effects of R^2 = .02 are interpreted as small, R^2 = .13 as medium, and R^2 = .26 as large effects.

Because of the explorative nature of our analyses, no adjustment for multiple testing has been conducted. Therefore, all results have to be interpreted as preliminary.

3. Results

First, the assumed still-face effect was investigated. Means and standard deviations of coded behavior can be seen in supplemental material S 1. The repeated measures ANOVA for *infant social positive engagement* was significant, F (2, 116) = 8.16, p <.001, η^2_p = .12. Bonferroni-corrected post-hoc tests revealed significant differences in the relative duration between play and still-face (M_{Diff} = 5.50, *95%-CI*[1.74, 9.25]) as well as between play and reunion (M_{Diff} = 3.47, *95%-CI*[0.14, 6.80]), however, no significant difference between still-face and reunion (M_{Diff} = -2.03, *95%-CI*[-5.08, 1.03]). *Infant social positive engagement* decreased from play to still-face, and increased from still-face to reunion, although, infants displayed less *infant social positive engagement* during the reunion compared to the play phase. There was a significant main effect for *infant protesting behavior*, F (2, 116) = 27.97, p <.001, η^2_p = .33, with significant post-hoc differences between play and still-face (M_{Diff} = -22.45, *95%-CI*[-34.08, -10.83]), play and reunion (M_{Diff} = -33.23, *95%-CI*[-45.31, -21.14]), and between still-face and

reunion (M_{Diff} = -10.77, *95%-CI*[-20.45, -1.10]). For *infant self-comfort (oral)*, no significant main effect could be found, F (2, 116) = 0.32, p = .728, η^2_p = .01, although infants displayed descriptively more oral self-comforting behavior in the still-face phase compared to play and reunion. For *infant self-comfort (self-clasp*), Greenhouse-Geisser correction was used to adjust for violations of sphericity. A significant main effect emerged, F (1.21, 70.42) = 13.82, p <.001, η^2_p = .19, with significant differences between play and still-face (M_{Diff} = -9.93, *95%-CI*[-16.94, -2.92]), as well as between still-face and reunion (M_{Diff} = 11.19, *95%-CI*[4.61, 17.76]), however not for play and reunion (M_{Diff} = 1.26, *95%-CI*[-1.33, 3.84]).

Results of the Spearman correlations are presented in the supplementary material (Table S 2). Regarding the block-wise hierarchical regression models, none of the first models were significant and none of the controlling factors significantly related to the criteria of interest (i.e., *infant protesting behavior*, *maternal positive affect*, *match*, *interactive reparation*, or *dyadic gaze*), except for *infant positive affect* that related to lower *infant gestational age* (see Table 3). Tables with model estimates and regression coefficients are only reported for models that could significantly explain more variance compared to the first model. The remaining tables can be found in the supplementary material (Tables S 3, S 4).

In the first block-wise hierarchical regression model, we tested for *infant positive affect*. The two factors HCC and SCC added significantly to the explanation of variance of the model with a significant change of R^2 = .15, F (2, 52) = 5.44, p = .007. *Infant positive affect* was significantly associated with lower *infant gestational age* and lower maternal *SCC* (see Table 3). Regarding the control variables, female infants exhibited more *infant positive* affect than male infants (see Table 3). The model coefficients and R^2 can be seen in Table 3.

Table 3

Results of the multiple regression models for infant positive affect.

	B	*SE*	*β*	95% CI	*t*	*p*	R^2
Model 1							**.16***
Intercept	242.65	75.32		[91.64, 393.66]	3.22	.002*	
Maternal age	-0.63	0.53	-0.16	[-1.70, 0.44]	-1.18	.241	

Parity	-5.57	4.60	-0.16	[-14.80, 3.66]	-1.21	.232	
Gestational age	-0.72	0.26	-0.35	[-1.24, -0.20]	-2.77	.008**	
Infant sex	-3.98	4.38	-0.12	[-12.76, 4.80]	-0.91	.367	
Model 2							**.31****
Intercept	325.00	74.96		[174.57, 475.42]	4.34	<.001***	
Maternal age	-0.75	0.51	-0.19	[-1.77, 0.27]	-1.47	.147	
Parity	-2.86	4.42	-0.08	[-11.72, 6.01]	-0.65	.521	
Gestational age	-0.83	0.24	-0.41	[-1.31, -0.34]	-3.41	.001**	
Infant sex	-7.27	4.19	0.21	[-15.68, 1.15]	-1.73	.089†	
HCC	11.18	10.11	-0.14	[-9.10, 31.46]	1.11	.274	
SCC	-24.92	8.25	-0.37	[-41.48, -8.36]	-3.02	.004**	

Note: N = 59; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed); ** $p < .01$, * $p < .05$, † $p < .10$.

For *infant protesting behavior* the second model could also explain significantly more variance compared to the first model, F (2, 52) = 4.94, p = .011. The change in R^2 was 15%. However, the overall model revealed only a statistical trend, with higher maternal *SCC* relating to *infant protesting behavior* and male infants displaying more *infant protesting behavior* compared to girls (see Table 4 for model coefficients and R^2).

Table 4

Results of the multiple regression models for infant protesting behavior*.*

	B	*SE*	*β*	95% CI	*t*	*p*	R^2
Model 1							**.05**
Intercept	-129.48	181.88		[-494.14, 235.17]	-0.71	.480	
Maternal age	0.23	1.29	0.03	[-2.35, 2.80]	0.18	.860	
Parity	6.75	11.12	0.09	[-15.54, 29.03]	0.61	.546	
Gestational age	0.56	0.63	0.12	[-0.69, 1.81]	0.90	.375	
Infant sex	14.83	10.58	0.19	[6.37, 36.03]	1.40	.167	
Model 2							**.21**†
Intercept	-313.18	182.48		[-679.36, 52.99]	-1.72	.092	
Maternal age	0.58	1.24	0.06	[-1.91,3.07]	0.47	.643	
Parity	-0.01	10.76	0.00	[-21.59, 21.58]	-0.00	*n.s.*	
Gestational age	0.81	0.59	0.18	[-0.37, 2.00]	1.37	.176	
Infant sex	22.62	10.21	0.29	[-2.14, 43.11]	2.22	.031*	
HCC	-31.56	24.60	-0.17	[-80.93, 17,81]	-1.28	.205	
SCC	55.62	20.08	0.36	[-15.32, 95.93]	2.77	.008**	

Note: N = 59; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC =

maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed); ** p < .01, † p <.10, *n.s.* p = 1.000.

Another block-wise hierarchical regression model tested for *maternal positive affect*. However, adding the two additional factors *HCC and SCC* did not explain significantly more variance, F (2, 52) = 0.09, p = .919, with a change in R^2 = .00. None of the factors were significant (see supplementary Table S 3 for model coefficient and R^2).

Testing for *match*, the second model could explain significantly more variance than the first model, F (2, 52) = 5.77, p = .005, with a change in R^2 of 16%. *Match* significantly related to lower maternal *SCC* (all model coefficients and R^2 can be seen in Table 5). Further, dyads with lower *gestational age* and female infants also exhibited more *match* (see Table 5).

Table 5

Results of the multiple regression models for match*.*

	B	*SE*	*β*	95% CI	*t*	*p*	R^2
Model 1							**.13**
Intercept	224.66	95.32		[33.55, 415.77]	2.36	.022*	
Maternal age	-0.64	0.67	-0.13	[-1.99, 0.71]	-0.95	.345	
Parity	-8.41	5.83	-0.20	[-20.09, 3.27]	-1.44	.155	
Gestational age	-0.62	0.33	-0.25	[-1.28, 0.04]	-1.89	.064†	
Infant sex	-7.06	5.54	-0.16	[-18.17, 4.05]	-1.27	.208	
Model 2							**.28****
Intercept	333.21	94.38		[143.83, 522.60]	3.53	<.001***	
Maternal age	-0.78	0.64	-0.16	[-2.06, 0.51]	-1.21	.231	
Parity	-5.03	5.56	-0.12	[-16.20, 6.13]	-0.91	.370	
Gestational age	-0.77	0.31	-0.31	[-1.38, -0.15]	-2.51	.015*	
Infant sex	-11.25	5.28	-0.26	[-21.85, -0.66]	-2.13	.038*	
HCC	12.85	12.72	0.13	[-12.69, 38.38]	1.01	.317	
SCC	-32.84	10.39	-0.39	[-53.69, -12.00]	-3.16	.003**	

Note: N = 59; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed); *** p <.001, ** p < .01, * p < .05, † p <.10.

For *interactive reparation*, the second model could not explain significantly more variance compared to the first model, F (2, 23) = 0.04, p = .965. The change of R^2 was 0%.

None of the factors were significant. All model coefficient and R^2 can be seen in the supplementary material (Table S 4).

The last block-wise hierarchical regression model tested for *dyadic eye-contact*. Adding the factors *HCC* and *SCC* to the model added significantly to the explanation of variance, F (2, 52) = 3.45, p = .039, with a change in R^2 = .11. The second model was not significant, with lower maternal *SCC* relating only trend-wise to *dyadic eye-contact* (see Table 6). There was also a trend that dyads with female infants showed more dyadic eye-contact compared to dyads with male infants (see Table 6). All model coefficients and R^2 can be seen in Table 6.

Table 6

Results of the multiple regression models for dyadic eye-contact*.*

	B	*SE*	*β*	95% CI	*t*	*p*	R^2
Model 1							**.09**
Intercept	179.14	101.90		[-0.25, 3.84]	1.76	.084†	
Maternal age	-0.59	0.72	-0.11	[-0.02, 0.01]	-0.82	.418	
Parity	-8.07	6.23	-0.18	[-0.21, 0.04]	-1.30	.201	
Gestational age	-0.44	0.35	-0.17	[-0.01, 0.00]	-1.26	.215	
Infant sex	-7.00	5.92	-0.16	[-0.19, 0.05]	-1.18	.243	
Model 2							**.19**†
Intercept	276.15	104.78		[0.66, 4.86]	2.64	.011*	
Maternal age	-0.65	0.71	-0.12	[-0.02, 0.01]	-0.91	.366	
Parity	-5.55	6.18	-0.12	[-0.18, 0.07]	-0.90	.373	
Gestational age	-0.57	0.34	-0.22	[-0.01, 0.00]	-1.68	.100	
Infant sex	-10.40	5.86	-0.23	[-0.22, 0.01]	-1.78	.082†	
HCC	6.77	1.134	0.06	[-0.22, 0.35]	0.48	.634	
SCC	-29.33	11.53	-0.33	[-0.53, -0.06]	-2.54	.014**	

Note: N = 59; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration, SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G); ** p < .01, * p < .05, † p <.10.

4. Discussion

The current study analyzed the association of maternal postpartum cortisol levels on mother-infant-interaction during a standardized FFSF using hair and saliva samples to measure situational and cumulated maternal cortisol levels. It could be shown that the standardized FFSF led to less positive and more negative affect after maternal non-

responsiveness, aligning with previous studies (Mesman et al., 2009). Regarding the main research question on whether maternal situational and cumulated cortisol levels, indicated by SCC and HCC, relate to different aspects of mother-infant-interaction, the analyses showed a significant association of SCC on infant positive affect, infant protesting behavior, dyadic match, and dyadic eye-contact during the reunion phase. The variance explained could be considered as small to moderate. HCC did not significantly relate to any of the investigated aspects of the mother-infant-interaction. Additionally, the analyses revealed significant associations with infant gestational age, and infant sex assigned at birth on infant positive affect, infant protest behavior and match. Additional analyses in the supplementary sample included two more factors, time of day and maternal depressive symptoms, that were entered as additional factors in the hierarchical regression model as both factors can potentially affect the association between SCC and HCC with different maternal, infant and dyadic behaviors. However, compared to the models presented in the main text, the additional factors time of day and maternal depressive symptoms were not significantly associated with the investigated aspects of the mother-infant-interaction. The current results should be considered as preliminary, especially, as the effect of situational cortisol possibly needs to be investigated under more standardized conditions. However, the results give valuable insights into the complex interplay between maternal cortisol and mother-infant-interaction and possible endpoints for future studies.

To investigate the association of maternal cortisol on the mother-infant-interaction, the current study used SCC as a marker for situational and HCC as a retrospective marker for cumulated maternal cortisol levels. We found a significant negative association between maternal SCC and infant positive affect, dyadic match, and dyadic eye-contact, as well as a positive association with infant protest during the reunion phase of the FFSF, while maternal positive affect and interactive reparation were unrelated to SCC. Mueller et al. (2021) demonstrated that offering an experimental stressor to the mother prior to the FFSF led to increased infant’s distress during the FFSF and that observations had to be terminated earlier,

although the authors did not control for an increase in maternal cortisol levels. Cortisol elevation following a stressor can demand cognitive resources that potentially impair maternal sensitivity, as indicated by studies that found effects of cortisol on cognitive flexibility (Goldfarb et al., 2017; Plessow et al., 2011) and on a shift from sensitive to more habitual behaviors (Schwabe & Wolf, 2013). This can possibly also affect the mother-infant-interaction, although this was not yet addressed in a corresponding study design. Thompson and Trevathan (2008) did not facilitate an experimental stressor, however, found that cortisol increase could be associated with less synchronized behavior between mothers and their 3-month-old infants during a free-play. This could reflect the aforementioned association between cortisol and maternal capabilities to engage in sensitive caregiving. The current study did not incorporate a standardized stress paradigm. Therefore, it is not clear if and what could have led to an increase in SCC in some mother, but not in others. The arrival to the laboratory in the current study is one factor that may has facilitated a cortisol response in some mothers more than others resulting in heightened SCC at the beginning of the experiment. An increase in cortisol may have resulted in impaired maternal capabilities to engage in dyadic behavior, which was reflected by less dyadic match and eye-contact in dyads with higher maternal SCC. Following the mutual regulation model (DiCorcia & Tronick, 2011; Gianino & Tronick, 1988), a mother with higher cortisol levels possibly had more difficulties to repair the interaction with her infant and to change back from a mismatching state to a matching state. This could be associated with cognitive and behavioral impairments due to momentary elevated cortisol levels (Goldfarb et al., 2017; Plessow et al., 2011; Schwabe & Wolf, 2013). A mere maternal positive affect, which was unrelated to SCC in the current study, possibly was insufficient to co-regulate infants after maternal unresponsiveness. Further, infants of mothers with higher SCC exhibited more negative and less positive affect during the reunion, possibly as an infant's response towards less dyadic coordination, which is comparable to the findings of Mueller et al. (2021). A reciprocity between maternal elevated cortisol levels and infant affect during interactions, however, cannot be ruled out either. Previous research suggests that infants born to mothers with increased cortisol levels during pregnancy are in general higher in irritability (Takegata et

al., 2021), which may contribute to less favorable dyadic and positive interaction patterns. Therefore, maternal cortisol responses could be generally higher due to a more difficult temperament of infants and vice versa regardless of the study setting. Other studies reported associations between averaged cortisol levels during a mother-infant-observation with maternal sensitivity (Finegood et al., 2016), or intrusive maternal behavior (Mills-Koonce et al., 2009), which is surprising as maternal positive affect and interactive reparation seemed unaffected by SCC in the current study. Otherwise, studies also linked higher cortisol levels measured from saliva to more adequate caregiving behavior and a higher alertness towards infant cues (Fleming et al., 1997). Shortly after birth, cortisol may lead to more caregiving behavior and higher responsivity towards infant cues (Bos, 2017), whereas later on, elevated levels of cortisol can be associated with less optimal caregiving, maternal sensitivity and responsiveness, and more negative interaction patterns (Mills-Koonce et al., 2009). Lastly, some studies also found no such effects of cortisol on maternal caregiving behavior (Bos et al., 2018). That the current study found no association between SCC and maternal behavior may depend on the study design itself. The laboratory observation may facilitate more positive behaviors in mothers (Belsky, 1980; Zegiob et al., 1975) as indicated by high amounts of maternal smiling and positive vocalizations during more than 90% of the time during the play and reunion phase of the current study. Studies found that mothers engage in a more positive (socially desirable) way in laboratory observations compared to more naturalistic at home observations (Belsky, 1980), and that being uninformed about being observed during a mother-infant-interaction led to less positive behavior and less adjusting of infant's behavior compared to being informed (Zegiob et al., 1975). However, at home observations potentially lack standardization and comparability, and leaving the women uninformed about being observed faces ethical and privacy policy considerations. Further, Mesman et al. (2013) found comparable results in a FFSF conducted at families homes, suggesting reliable still-face effects regardless of the setting.

In the current study, we were also interested in the association between more long-time cortisol levels and the mother-infant-interaction using HCC, a marker for cumulated cortisol levels. No significant association with maternal, infant, or dyadic behavior emerged. This contradicts the results of Tarullo et al. (2017) that higher HCC related to more negative and fewer positive behaviors in mother-infant-dyads during a free play interaction. Other studies obtained maternal HCC during pregnancy. They found that elevated HCC in the third trimester could be associated with more intrusive maternal behavior as well as maternal withdrawal during the FFSF paradigm, however, only when mothers additionally exhibited depressive symptoms at 4 months postpartum (Khoury et al., 2020). Nystrom-Hansen et al. (2019) assessed maternal HCC during the third trimester and again 4 months postpartum and found a significant association between maternal HCC and mental illnesses (e.g., depression, bipolar disorder, schizophrenia) as well as disrupted mother-infant-interaction 4 months postpartum. The authors further found HCC (during the third trimester and 4 months postpartum) to mediate the association between maternal mental illness and disrupted mother-infant-interaction (Nystrom-Hansen et al., 2019). In the current study, maternal mental illness was not considered as an associated factor. Conversely, long-term mental illnesses can also downregulate the HPA axis and lower cortisol levels (Pochigaeva et al., 2017). Our sample had relatively low maternal HCC compared to other studies (Kirschbaum et al., 2009), possibly due to only mild stressors during the past months, a chronic downregulation of the HPA axis, or hormonal changes associated with pregnancy and birth, which cannot further be assessed. An additional control for mental illnesses or life events is indicated, however, self-reported measures for mental illnesses and HCC seem to not always be correlated (Braig et al., 2016). In the current study, maternal depressive symptoms assessed by the EPDS did not change the overall results Further, a preceded pregnancy and concomitant tremendous changes in hormonal levels potentially impair the interpretation of cortisol levels. HCC as a retrospective marker for cumulative cortisol levels reflects a certain period of time depending on the length of the observed hair strand. As infants were 4 months of age, we investigated a time frame shortly after birth which may be specifically affected by the hormonal changes due to

pregnancy and birth: Hair cortisol levels of pregnant women in contrast to non-pregnant women are increased from third trimester up to 3 months postpartum (Kirschbaum et al., 2009). Galbally et al. (2019) found HCC increases during pregnancy until 3 months postpartum and then decrease from 3 months postpartum to 12 months postpartum. The current study might have captured a sensitive period with rapid changes in maternal cortisol levels due to hormonal shifts, and maternal HCC may reflect these changes rather than the presumed cumulated cortisol levels as a consequence of stress. Another time point when hormonal levels of mothers are comparable to levels of non-pregnant women may be indicated. However, more research is highly needed as current results are scarce and only few studies investigated HCC during the postpartum period.

The current study considered maternal and infant characteristics that were described to be important while investigating associations between maternal cortisol levels and the mother-infant-interaction. However, only infant gestational age and sex assigned at birth significantly related to infant affect and dyadic interaction. Infant gestational age was negatively associated with infant positive engagement and dyadic match during reunion, challenging a broad body of literature suggesting that lower gestational age, often reflected by preterm infants, relates to less positive and synchronized parent-infant-interaction (for an overview see e.g., Bilgin & Wolke, 2015). As the current study investigated only full-term infants (i.e., born ≥ 37^{th} week of gestation) applicability of these results is unclear. Research on the effects of gestational age within the normal range is scarce, suggesting the need for more comprehensive studies on gestational age within this range. Our analysis revealed that female infants displayed more positive affect, less protest behavior, as well as more dyadic interaction during the reunion phase. Findings considering sex differences in infant affect during the FFSF are heterogeneous (Alexander & Wilcox, 2012), however, there are studies reporting that infant boys seem to be more prone to the maternal still-face (Weinberg et al., 1999), which can also be inferred from our study results. As sex differences in infant behavior are evident quite early after birth and a reinforcement of such gender differences by socialization factors is likely

(Alexander & Wilcox, 2012), infant sex assigned at birth seems to be an important variable to include in studies investigating mother-infant-interactions.

The current study has some methodological considerations and limitations. HCC can be affected by several other factors, e.g., socioeconomic status, traumatic events, UV radiation, chemical hair treatment or washing (Greff et al., 2019; Stalder et al., 2017), that may impede study results. Further, cortisol measured in hair or saliva is directly affected by sampling time, physical activity, food intake, sleep, or smoking (Kirschbaum & Hellhammer, 2008; Stalder et al., 2017), for which the current study did not account and should therefore be interpreted as preliminary. However, as can be seen by the additional analyses in the supplementary material, the addition of time of sampling of saliva samples as a factor did not change the overall results. The findings of this study are correlational in nature, and thus, caution must be exercised in drawing causal inferences. To improve the current correlative study design, future studies could integrate standardized stress paradigms (see Mueller et al., 2021) and endocrine markers that reflect the human stress response to test for effects on the mother-infant-interaction. Next, although the laboratory assessment allows for standardization, it may not capture the mother-infant-interaction in a naturalistic way (Belsky, 1980; Zegiob et al., 1975) as noted earlier discussing the results on SCC. As our sample was quite homogenous and high in educational levels, a more diverse sample is indicated. The current study only investigated mothers, however, for infant development other caregivers like fathers seem just as important (Jansen et al., 2024) and should be evenly investigated in future studies.

Lastly, we want to emphasize some strengths of the current study. To the best of our knowledge, this is one of the first studies investigating the effects of maternal cortisol levels and the mother-infant-interaction during a standardized FFSF using two endocrine parameters HCC and SCC, alongside the study by Tarullo et al. (2017) who, however, used a free play observation. Our study adds valuable information about the impact of maternal cortisol during the early postpartum period for the mother-infant-interaction while controlling for several

mother and infant characteristics. Further, to analyze the mother-infant-interaction, we have used a microanalytical approach. This allows to capture more fine-grained patterns of interaction compared to macro-analytical approaches (Lotzin et al., 2015).

5. Conclusion

This is one of the first studies examining the association between maternal cortisol and the mother-infant-interaction by employing maternal HCC and SCC in one study under consideration of potentially relevant maternal and infant characteristics. Model estimates have to be considered as low to moderate, but the results indicated that maternal SCC is related to infant and dyadic behavior, while HCC did not relate to any facet of the mother-infant-interaction. Interestingly, maternal positive interaction did not reveal any association with the considered variables, either because the behavior is rather unaffected by maternal cortisol, or it was distorted by the laboratory setting. Infant gestational age and sex assigned at birth emerged as significant factors relating to mother-infant-interaction, suggesting the need for their inclusion as control variables in future studies. Given the exploratory approach of our study, our results need to be validated in future research that allows for more causal interpretation.

Investigating the effects of maternal cortisol during the postpartum period on the mother-infant-interaction is of high significance, as dyads that cannot effectively repair their interaction after a disruption could face risks: infants potentially fail to learn adequate self-regulation strategies (DiCorcia & Tronick, 2011), and caregivers potentially build negative expectations when interacting with their infants (Seymour et al., 2015), possibly leading to long-term psychosocial consequences and emotion-regulation problems (Deans, 2020). Therefore, it is crucial that future research, building on the preliminary findings presented here, be conducted to identify caregivers at risk for developing problematic parent-infant interactions during the early stages of infant development.

Funding

The authors received no specific funding for this work.

Author statement

The manuscript is not under consideration for publication elsewhere. The work has not been previously published. The manuscript was approved by all authors.

Declaration of competing interest

The authors declare that there is no competing interest.

CRediT authorship contribution statement

Luisa Ernsten: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – Original Draft, Visualization

Sabine Seehagen: Methodology, Writing – Review & Editing

Nora Nonnenmacher: Methodology, Writing – Review & Editing

Juliane Tautz: Methodology, Writing – Review & Editing

Oliver T. Wolf: Methodology, Writing – Review & Editing

Martin Heil: Methodology, Writing – Review & Editing

Nora K. Schaal: Conceptualization, Methodology, Writing – Review & Editing, Supervision

6. References

Alexander, G. M., & Wilcox, T. (2012). Sex Differences in Early Infancy. *Child Development Perspectives*, *6*(4), 400-406. https://doi.org/https://doi.org/10.1111/j.1750-8606.2012.00247.x

Almanza-Sepulveda, M. L., Fleming, A. S., & Jonas, W. (2020). Mothering revisited: A role for cortisol? *Horm Behav*, *121*, 104679. https://doi.org/10.1016/j.yhbeh.2020.104679

Belsky, J. (1980). MOTHER-INFANT INTERACTION AT HOME AND IN THE LABORATORY - A COMPARATIVE-STUDY. *Journal of Genetic Psychology*, *137*(1), 37-47. https://doi.org/10.1080/00221325.1980.10532800

Bergant, A. M., Nguyen, T., Heim, K., Ulmer, H., & Dapunt, O. (1998). German validation of the Edinburgh Postpartum Depression Scale. *Deutsche Medizinische Wochenschrift*, *123*, 35-40. https://doi.org/10.1055/s-2007-1023895

Bhake, R. C., Kluckner, V., Stassen, H., Russell, G. M., Leendertz, J., Stevens, K., Linthorst, A. C. E., & Lightman, S. L. (2019). Continuous Free Cortisol Profiles—Circadian Rhythms in Healthy Men. *The Journal of Clinical Endocrinology & Metabolism*, *104*(12), 5935-5947. https://doi.org/10.1210/jc.2019-00449

Bilgin, A., & Wolke, D. (2015). Maternal Sensitivity in Parenting Preterm Children: A Meta-analysis. *Pediatrics*, *136*(1), e177-193. https://doi.org/10.1542/peds.2014-3570

Blair, C., & Ku, S. (2022). A Hierarchical Integrated Model of Self-Regulation [Conceptual Analysis]. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.725828

Booth, A. T., Macdonald, J. A., & Youssef, G. J. (2018). Contextual stress and maternal sensitivity: A meta-analytic review of stress associations with the Maternal Behavior Q-Sort in observational studies. *Developmental Review*, *48*, 145-177. https://doi.org/https://doi.org/10.1016/j.dr.2018.02.002

Bos, P. A. (2017). The endocrinology of human caregiving and its intergenerational transmission. *Dev Psychopathol*, *29*(3), 971-999. https://doi.org/10.1017/s0954579416000973

Bos, P. A., Hechler, C., Beijers, R., Shinohara, K., Esposito, G., & de Weerth, C. (2018). Prenatal and postnatal cortisol and testosterone are related to parental caregiving quality in fathers, but not in mothers. *Psychoneuroendocrinology*, *97*, 94-103. https://doi.org/10.1016/j.psyneuen.2018.07.013

Bowlby, J., & Ainsworth, M. (2000). The Origins of Attachment Theory. In S. Goldberg, R. Muir, & J. Kerr (Eds.), *Attachment Theory. Social Developmental, and Clinical Perspectives* (Vol. 1, pp. 45-84). The Analytic Press, Inc.

Braig, S., Grabher, F., Ntomchukwu, C., Reister, F., Stalder, T., Kirschbaum, C., Rothenbacher, D., & Genuneit, J. (2016). The Association of Hair Cortisol with Self-Reported Chronic Psychosocial Stress and Symptoms of Anxiety and Depression in Women Shortly after Delivery. *Paediatric and Perinatal Epidemiology*, *30*(2), 97-104. https://doi.org/https://doi.org/10.1111/ppe.12255

Braungart-Rieker, J. M., Zentall, S., Lickenbrock, D. M., Ekas, N. V., Oshio, T., & Planalp, E. (2014). Attachment in the making: mother and father sensitivity and infants' responses during the Still-Face Paradigm. *J Exp Child Psychol*, *125*, 63-84. https://doi.org/10.1016/j.jecp.2014.02.007

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Lawrence Erlbaum Associates.

Cole, L., & Kramer, P. R. (2016). Chapter 3.1 - The Endocrine System. In L. Cole & P. R. Kramer (Eds.), *Human Physiology, Biochemistry and Basic Medicine* (pp. 69-77). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-803699-0.00003-7

Cox, J. (2019). Thirty years with the Edinburgh Postnatal Depression Scale: voices from the past and recommendations for the future. *The British Journal of Psychiatry*, *214*(3), 127-129. https://doi.org/10.1192/bjp.2018.245

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry*, *150*, 782-786. https://doi.org/10.1192/bjp.150.6.782

Deans, C. L. (2020). Maternal sensitivity, its relationship with child outcomes, and interventions that address it: a systematic literature review. *Early Child Development and Care*, *190*(2), 252-275. https://doi.org/10.1080/03004430.2018.1465415

DiCorcia, J. A., & Tronick, E. (2011). Quotidian resilience: Exploring mechanisms that drive resilience from a perspective of everyday stress and coping. *Neuroscience & Biobehavioral Reviews*, *35*(7), 1593-1602. https://doi.org/https://doi.org/10.1016/j.neubiorev.2011.04.008

Emmanuel, E., & St John, W. (2010). Maternal distress: a concept analysis. *Journal of Advanced Nursing*, *66*(9), 2104-2115. https://doi.org/https://doi.org/10.1111/j.1365-2648.2010.05371.x

Feldman, R., Granat, A., Pariente, C., Kanety, H., Kuint, J., & Gilboa-Schechtman, E. (2009). Maternal depression and anxiety across the postpartum year and infant social engagement, fear regulation, and stress reactivity. *J Am Acad Child Adolesc Psychiatry*, *48*(9), 919-927. https://doi.org/10.1097/CHI.0b013e3181b21651

Finegood, E. D., Blair, C., Granger, D. A., Hibel, L. C., & Mills-Koonce, R. (2016). Psychobiological influences on maternal sensitivity in the context of adversity. *Dev Psychol*, *52*(7), 1073-1087. https://doi.org/10.1037/dev0000123

Fleming, A. S., Steiner, M., & Corter, C. (1997). Cortisol, hedonics, and maternal responsiveness in human mothers. *Horm Behav*, *32*(2), 85-98. https://doi.org/10.1006/hbeh.1997.1407

Galbally, M., van Rossum, E. F. C., Watson, S. J., de Kloet, E. R., & Lewis, A. J. (2019). Trans-generational stress regulation: Mother-infant cortisol and maternal mental health across the perinatal period. *Psychoneuroendocrinology*, *109*, 104374. https://doi.org/10.1016/j.psyneuen.2019.104374

Gianino, A., & Tronick, E. Z. (1988). The mutual regulation model: The infant's self and interactive regulation and coping and defensive capacities. In *Stress and coping across development.* (pp. 47-68). Lawrence Erlbaum Associates, Inc.

Goldfarb, E. V., Froböse, M. I., Cools, R., & Phelps, E. A. (2017). Stress and Cognitive Flexibility: Cortisol Increases Are Associated with Enhanced Updating but Impaired Switching. *Journal of Cognitive Neuroscience*, *29*(1), 14-24. https://doi.org/10.1162/jocn_a_01029

Greff, M. J. E., Levine, J. M., Abuzgaia, A. M., Elzagallaai, A. A., Rieder, M. J., & van Uum, S. H. M. (2019). Hair cortisol analysis: An update on methodological considerations and clinical applications. *Clinical Biochemistry*, *63*, 1-9. https://doi.org/https://doi.org/10.1016/j.clinbiochem.2018.09.010

Hammen, C. (2005). Stress and Depression. *Annual Review of Clinical Psychology*, *1*(1), 293-319. https://doi.org/10.1146/annurev.clinpsy.1.102803.143938

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, *34*(2), 163-171. https://doi.org/https://doi.org/10.1016/j.psyneuen.2008.10.026

IBL Internation GmbH. Arbeitsanleitung Cortisol Saliva Luminescence Immunoassay: Katalognummer RE62111/RE62119 [Version 2015-12].

Ilyka, D., Johnson, M. H., & Lloyd-Fox, S. (2021). Infant social interactions and brain development: A systematic review. *Neuroscience & Biobehavioral Reviews*, *130*, 448-469. https://doi.org/https://doi.org/10.1016/j.neubiorev.2021.09.001

Jansen, E., Marceau, K., Sellers, R., Chen, T., Garfield, C. F., Leve, L. D., Neiderhiser, J. M., Spotts, E. L., Roary, M., & Outcomes, p. c. f. E. i. o. C. H. (2024). The role of fathers in child development from preconception to postnatal influences: Opportunities for the National Institutes of Health Environmental influences on Child Health Outcomes (ECHO) program. *Developmental Psychobiology*, *66*(2), e22451. https://doi.org/https://doi.org/10.1002/dev.22451

Khoury, J. E., Bosquet Enlow, M., Patwa, M. C., & Lyons-Ruth, K. (2020). Hair cortisol in pregnancy interacts with maternal depressive symptoms to predict maternal disrupted interaction with her infant at 4 months. *Dev Psychobiol*, *62*(6), 768-782. https://doi.org/10.1002/dev.21950

Kirschbaum, C., & Hellhammer, D. H. (2008). Salivary Cortisol in Psychobiological Research: An Overview. *Neuropsychobiology*, *22*(3), 150-169. https://doi.org/10.1159/000118611

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (2008). The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, *28*(1-2), 76-81. https://doi.org/10.1159/000119004

Kirschbaum, C., Tietze, A., Skoluda, N., & Dettenborn, L. (2009). Hair as a retrospective calendar of cortisol production-Increased cortisol incorporation into hair in the third trimester of pregnancy. *Psychoneuroendocrinology*, *34*(1), 32-37. https://doi.org/10.1016/j.psyneuen.2008.08.024

Laurent, H. K., Harold, G. T., Leve, L., Shelton, K. H., & Van Goozen, S. H. M. (2016). Understanding the unfolding of stress regulation in infants. *Development and Psychopathology*, *28*(4pt2), 1431-1440. https://doi.org/10.1017/S0954579416000171

Lotzin, A., Lu, X., Kriston, L., Schiborr, J., Musal, T., Romer, G., & Ramsauer, B. (2015). Observational Tools for Measuring Parent–Infant Interaction: A Systematic Review. *Clinical Child and Family Psychology Review*, *18*(2), 99-132. https://doi.org/10.1007/s10567-015-0180-z

McEwen, B. S. (2004). Protection and Damage from Acute and Chronic Stress: Allostasis and Allostatic Overload and Relevance to the Pathophysiology of Psychiatric Disorders. *Annals of the New York Academy of Sciences*, *1032*(1), 1-7. https://doi.org/https://doi.org/10.1196/annals.1314.001

Mesman, J., Linting, M., Joosen, K. J., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2013). Robust patterns and individual variations: stability and predictors of infant behavior in the still-face paradigm. *Infant Behav Dev*, *36*(4), 587-598. https://doi.org/10.1016/j.infbeh.2013.06.004

Mesman, J., van Ijzendoorn, M., & Bakermans-Kranenburg, M. (2009). The many faces of the Still-Face Paradigm: A review and meta-analysis. *Developmental Review*, *30*, 120-162. https://doi.org/10.1016/j.dr.2009.02.001

Meyer, J., Novak, M., Hamel, A., & Rosenberg, K. (2014). Extraction and analysis of cortisol from human and monkey hair. *J Vis Exp*(83), e50882. https://doi.org/10.3791/50882

Meyer, J. S., & Novak, M. A. (2012). Minireview: Hair cortisol: a novel biomarker of hypothalamic-pituitary-adrenocortical activity. *Endocrinology*, *153*(9), 4120-4127. https://doi.org/10.1210/en.2012-1226

Mills-Koonce, W. R., Propper, C., Gariepy, J.-L., Barnett, M., Moore, G. A., Calkins, S., & Cox, M. J. (2009). Psychophysiological correlates of parenting behavior in mothers of young children. *Developmental Psychobiology*, *51*(8), 650-661. https://doi.org/https://doi.org/10.1002/dev.20400

Mueller, I., Snidman, N., DiCorcia, J. A., & Tronick, E. (2021). Acute Maternal Stress Disrupts Infant Regulation of the Autonomic Nervous System and Behavior: A CASP Study [Original Research]. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.714664

Müller, M., Zietlow, A.-L., Klauser, N., Woll, C., Nonnenmacher, N., Tronick, E., & Reck, C. (2022). From Early Micro-Temporal Interaction Patterns to Child Cortisol Levels: Toward the Role of Interactive Reparation and Infant Attachment in a Longitudinal Study [Original Research]. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.807157

Nystrom-Hansen, M., Andersen, M. S., Khoury, J. E., Davidsen, K., Gumley, A., Lyons-Ruth, K., MacBeth, A., & Harder, S. (2019). Hair cortisol in the perinatal period mediates associations between maternal adversity and disrupted maternal interaction in early infancy [Article]. *Developmental Psychobiology*, *61*(4), 543-556. https://doi.org/10.1002/dev.21833

Plessow, F., Fischer, R., Kirschbaum, C., & Goschke, T. (2011). Inflexibly Focused under Stress: Acute Psychosocial Stress Increases Shielding of Action Goals at the Expense of Reduced Cognitive Flexibility with Increasing Time Lag to the Stressor. *Journal of Cognitive Neuroscience*, *23*(11), 3218-3227. https://doi.org/10.1162/jocn_a_00024

Pochigaeva, K., Druzhkova, T., Yakovlev, A., Onufriev, M., Grishkina, M., Chepelev, A., Guekht, A., & Gulyaeva, N. (2017). Hair cortisol as a marker of hypothalamic-pituitary-adrenal Axis activity in female patients with major depressive disorder. *Metabolic Brain Disease*, *32*(2), 577-583. https://doi.org/10.1007/s11011-017-9952-0

Pruessner, J. C., & Ali, N. (2015). Neuroendocrine Mechanisms of Stress Regulation in Humans. In *Neuroendocrinology of Stress* (pp. 121-142). https://doi.org/https://doi.org/10.1002/9781118921692.ch6

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, *28*(7), 916-931. https://doi.org/10.1016/s0306-4530(02)00108-7

Ranjit, N., Young, E. A., Raghunathan, T. E., & Kaplan, G. A. (2005). Modeling cortisol rhythms in a population-based study. *Psychoneuroendocrinology*, *30*(7), 615-624. https://doi.org/https://doi.org/10.1016/j.psyneuen.2005.02.003

Reck, C., Noe, D., & Cenciotti, F. (2009). Infant and Caregiver Engagement Phases (ICEP): Heidelberg Version. In U. o. Heidelberg (Ed.). Heidelberg, Germany.

Schwabe, L., & Wolf, O. T. (2013). Stress and multiple memory systems: from 'thinking' to 'doing'. *Trends Cogn Sci*, *17*(2), 60-68. https://doi.org/10.1016/j.tics.2012.12.001

Seymour, M., Giallo, R., Cooklin, A., & Dunning, M. (2015). Maternal anxiety, risk factors and parenting in the first post-natal year. *Child: Care, Health and Development*, *41*(2), 314-323. https://doi.org/https://doi.org/10.1111/cch.12178

Shin, H., Park, Y.-J., Ryu, H., & Seomun, G.-A. (2008). Maternal sensitivity: a concept analysis. *Journal of Advanced Nursing*, *64*(3), 304-314. https://doi.org/https://doi.org/10.1111/j.1365-2648.2008.04814.x

Short, S. J., Stalder, T., Marceau, K., Entringer, S., Moog, N. K., Shirtcliff, E. A., Wadhwa, P. D., & Buss, C. (2016). Correspondence between hair cortisol concentrations and 30-day integrated daily salivary and weekly urinary cortisol measures.

Psychoneuroendocrinology, *71*, 12-18. https://doi.org/https://doi.org/10.1016/j.psyneuen.2016.05.007

Singh Solorzano, C., Serwinski, B., Grano, C., & Steptoe, A. (2023). Longitudinal association between saliva and hair cortisol concentration: A systematic comparison. *Psychoneuroendocrinology*, *156*, 106340. https://doi.org/https://doi.org/10.1016/j.psyneuen.2023.106340

Sjögren, E., Leanderson, P., & Kristenson, M. (2006). Diurnal saliva cortisol levels and relations to psychosocial factors in a population sample of middle-aged swedish men and women. *International Journal of Behavioral Medicine*, *13*(3), 193-200. https://doi.org/10.1207/s15327558ijbm1303_2

Smith, S. M., & Vale, W. W. (2006). The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress. *Dialogues in Clinical Neuroscience*, *8*(4), 383-395. https://doi.org/10.31887/DCNS.2006.8.4/ssmith

Stalder, T., & Kirschbaum, C. (2012). Analysis of cortisol in hair – State of the art and future directions. *Brain, Behavior, and Immunity*, *26*(7), 1019-1029. https://doi.org/https://doi.org/10.1016/j.bbi.2012.02.002

Stalder, T., Steudte-Schmiedgen, S., Alexander, N., Klucken, T., Vater, A., Wichmann, S., Kirschbaum, C., & Miller, R. (2017). Stress-related and basic determinants of hair cortisol in humans: A meta-analysis. *Psychoneuroendocrinology*, *77*, 261-274. https://doi.org/https://doi.org/10.1016/j.psyneuen.2016.12.017

Stanley, C., Murray, L., & Stein, A. (2004). The effect of postnatal depression on mother-infant interaction, infant response to the still-face perturbation, and performance on an instrumental learning task. *Development and psychopathology*, *16*, 1-18. https://doi.org/10.10170S0954579404040386

Takegata, M., Matsunaga, A., Ohashi, Y., Toizumi, M., Yoshida, L. M., & Kitamura, T. (2021). Prenatal and Intrapartum Factors Associated With Infant Temperament: A Systematic Review [Systematic Review]. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.609020

Tarullo, A. R., St John, A. M., & Meyer, J. S. (2017). Chronic stress in the mother-infant dyad: Maternal hair cortisol, infant salivary cortisol and interactional synchrony. *Infant behavior & development*, *47*, 92-102. https://doi.org/10.1016/j.infbeh.2017.03.007

Tester-Jones, M., Karl, A., Watkins, E., & O'Mahen, H. (2016). Rumination in dysphoric mothers negatively affects mother–infant interactions. *Journal of Child Psychology and Psychiatry*, *58*. https://doi.org/10.1111/jcpp.12633

Thompson, L. A., & Trevathan, W. R. (2008). Cortisol reactivity, maternal sensitivity, and learning in 3-month-old infants. *Infant Behavior & Development*, *31*(1), 92-106. https://doi.org/10.1016/j.infbeh.2007.07.007

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *J Am Acad Child Psychiatry*, *17*(1), 1-13. https://doi.org/10.1016/s0002-7138(09)62273-1

Tronick, E., Mueller, I., DiCorcia, J., Hunter, R., & Snidman, N. (2021). A Caretaker Acute Stress Paradigm: Effects on behavior and physiology of caretaker and infant. *Developmental Psychobiology*, *63*(2), 237-246. https://doi.org/https://doi.org/10.1002/dev.21974

Tsigos, C., Kyrou, I., Kassi, E., & Chrousos, G. P. (2020). *Stress: Endocrine Physiology and Pathophysiology* https://www.ncbi.nlm.nih.gov/books/NBK278995/

Weinberg, M., Tronick, E., Cohn, J., & Olson, K. (1999). Gender differences in emotional expressivity and self-regulation during early infancy. *Developmental Psychology*, *35*, 175-188. https://doi.org/10.1037//0012-1649.35.1.175

Zegiob, L. E., Arnold, S., & Forehand, R. (1975). An Examination of Observer Effects in Parent-Child Interactions. *Child Development*, *46*(2), 509-512. https://doi.org/10.2307/1128149

Zhang, Q., Chen, Z., Chen, S., Yu, T., Wang, J., Wang, W., & Deng, H. (2018). Correlations of hair level with salivary level in cortisol and cortisone. *Life Sciences*, *193*, 57-63. https://doi.org/https://doi.org/10.1016/j.lfs.2017.11.037

Table S 1

Means and standard deviations of coded behavior and codes used as criteria.

	Play		Still-Face		Reunion	
	M	*SD*	*M*	*SD*	*M*	*SD*
Infant social positive engagement	9.51	10.76	4.01	7.53	6.03	7.54
Infant protesting behavior	11.71	20.42	34.16	34.05	44.93	39.76
Infant self-comforting (oral)	12.98	21.81	15.04	21.85	13.05	22.23
Infant self-comforting (self-clasp)	10.34	14.59	20.26	26.68	9.08	15.60
Infant positive affect	37.20	23.66	32.85	28.16	22.79	21.94
Maternal positive affect	93.98	10.58	-	-	91.30	10.46
Match	36.43	23.47	-	-	22.14	21.66
Interactive reparation	7.46	17.12	-	-	14.42	29.72
Dyadic eye-contact	38.69	23.22	-	-	28.79	22.67

Note: N = 59; all codes are reported as relative duration in percent, except for *interactive reparation* which is reported in seconds; as mothers were instructed to pause any (dyadic) behavior, *maternal positive affect*, *match*, *interactive reparation*, and *dyadic eye-contact* during the still-face phase are not reported.

Table S 2

Spearman correlations of study variables.

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1.	Maternal age	-										
2.	Parity	.28*	-									
3.	Gestational age	-.13	-.23†	-								
4.	Infant sex	-.15	-.05	.05	-							
5.	HCC	.14	-.21	.08	.16	-						
6.	SCC	.12	.20	-.19	-.26*	-.13	-					
7.	Infant positive engagement	-.11	-.06	-.18	-.18	.15	-.33*	-				
8.	Infant protest behavior	.07	.05	.07	.15	-.13	.31*	-.72***	-			
9.	Maternal positive engagement	-.12	-.21	.10	-.19	-.08	-.01	.33*	-.27*	-		
10.	Match	-.12	-.14	-.12	-.21	.15	-.35**	.97***	-.74***	.38**	-	
11.	Interactive reparation	.06	.06	.35†	-.02	-.01	-.09	-.26	.41*	-.28	-.16	
12.	Dyadic gaze	-.06	-.17	-.03	-.18	.02	-.28*	.77***	-.42**	.39**	.80***	-.04

Note: N = 59 (except for *interactive reparation*, *n* = 30, as 29 dyads did not show any interactive reparation during reunion); parity is dummy coded as 0 = primiparous and 1 = multiparous, sex is dummy coded as 0 = female and 1 = male, HCC = hair cortisol concentrations (log-transformed), SCC = salivary cortisol concentrations indicated as log-transformed area under the curve with respect to ground (AUC_G); *** p <.001, ** p < .01, * p < .05, † p <.10.

Table S 3

Results of the multiple regression models for maternal positive affect.

	B	*SE*	*β*	95% CI	*t*	*p*	R^2
Model 1							**.08**
Intercept	58.37	47.12		[-36.10, 152.84]	1.24	.221	
Maternal age	0.03	0.33	0.01	[-0.64, 0.70]	0.08	.933	
Parity	-3.43	2.88	-0.17	[-9.21, 2.34]	-1.19	.238	
Gestational age	0.13	0.16	0.11	[-0.20, 0.45]	0.80	.429	
Infant sex	-4.25	2.74	-0.21	[-9.74, 1.24]	-1.55	.127	
Model 2							**.09**
Intercept	54.96	51.48		[-48.35, 158.27]	1.07	.291	
Maternal age	-0.01	0.35	-0.00	[-0.71, 0.70]	-0.02	.986	
Parity	-3.22	3.03	-0.16	[-9.31, 2.87]	-1.06	.294	
Gestational age	0.13	0.17	0.11	[-0.20, 0.47]	0.79	.431	
Infant sex	-4.34	2.88	-0.21	[-10.12, 1.44]	-1.51	.138	
HCC	2.65	6.94	0.05	[-11.28, 16.58]	0.38	.704	
SCC	1.02	5.67	0.03	[-10.36, 12.39]	0.18	.859	

Note: N = 59; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed).

Table S 4

Results of the multiple regression models for *interactive reparation*.

	B	*SE*	*β*	95% CI	*t*	*p*	R^2
Model 1							**.13**
Intercept	-236.89	186.53		[-621.07, 147.28]	-1.27	.216	
Maternal age	0.68	1.55	0.08	[-2.50, 3.87]	0.44	.662	
Parity	16.61	11.40	0.27	[-6.87, 40.09]	1.46	.157	
Gestational age	0.78	0.64	0.23	[-0.53, 2.09]	1.22	.233	
Infant sex	2.04	11.05	0.04	[-20.73, 24.81]	0.18	.855	
Model 2							**.13**
Intercept	-242.02	205.93		[-668.02, 183.98]	-1.18	.252	
Maternal age	0.47	1.92	0.06	[-3.50, 4.44]	0.25	.809	
Parity	16.08	12.45	0.27	[-9.68, 41.84]	1.29	.209	
Gestational age	0.77	0.67	0.23	[-0.61, 2.16]	1.15	.261	
Infant sex	-0.97	14.93	0.02	[-29.91, 31.85]	0.07	.949	
HCC	7.04	42.23	0.05	[-80.32, 94.41]	0.17	.869	
SCC	-5.44	25.83	0.04	[-48.00, 58.88]	0.21	.835	

Note: N = 48 (*n* = 11 dyads displayed no interactive reparation); parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration, SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G).

Table S 5

Pearson correlations of study variables including the additional factors day of time and the EDS score.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Infant positive affect	-												
2. Infant protest behavior	-.72***	-											
3. Maternal positive affect	.33*	-.27*	-										
4. Match	.97***	-.74***	.38**	-									
5. Dyadic eye-contact	.77***	-.42*	.39**	.80***	-								
6. Interactive reparation	-.26	.41*	-.28	-.16	-.04	-							
7. Maternal age	-.11	.07	-.12	-.12	-.06	.06	-						
8. Parity	-.06	.05	-.21	-.14	-.17	.06	.28*	-					
9. Gestational age	-.18	.07	.10	-.12	-.03	.35†	-.13	-.23†	-				
10. Infant sex	-.18	.15	-.19	-.21	-.18	-.02	-.15	-.05	.05	-			
11. HCC	.15	-.13	-.08	.15	.02	-.01	.14	-.21	.08	.16	-		
12. SCC	-.33*	.31*	-.01	-.35**	-.28*	-.09	.12	.20	-.19	-.26*	-.13	-	
13. Time of day	.22†	-.19	.08	.23†	.27*	.42*	-.05	-.15	-.04	.15	.03	-.25†	-
14. EPDS score	-.23†	.09	.06	-.19	-.16	.11	-.25†	-.05	-.22	-.06	-.16	-.04	-.04

Note: N = 59 (except for interactive reparation *n* = 30, as 29 dyads did not show any interactive reparation during reunion, and for EPDS score *n* = 57 as two scores were missing); parity is dummy coded as 0 = primiparous and 1 = multiparous, sex is dummy coded as 0 = female and 1 = male, HCC = hair cortisol concentrations (log-transformed), SCC = salivary cortisol concentrations indicated as log-transformed area under the curve with respect to ground (AUC_G), time of day dummy coded as '0 = before 12 pm' and '1 = after 12 pm'; ** $p < .01$, * $p < .05$, † $p < .10$ (two-tailed).

Table S 6

Results of the multiple regression models for infant positive affect *including the additional factors day of time and the EPDS score.*

	B	*SE*	β	95% CI lower	95% CI upper	*t*	*p*	R^2
Model 1								**.16†**
Intercept	251.42	78.21		94.47	408.36	3.21	.002**	
Maternal age	-0.71	0.55	-.17	-1.82	0.40	-1.29	.205	
Parity	-5.77	4.55	-.17	-14.89	3.35	-1.27	.210	
Gestational age	-0.74	0.26	-.38	-1.27	-0.21	-2.82	.007**	
Infant sex	-2.44	4.36	-.07	-11.18	6.31	-0.56	.579	
Model 2								**.37****
Intercept	297.91	78.19		140.70	455.13	3.81	<.001	
Maternal age	-0.89	0.52	-.21	-1.93	0.16	-1.70	.096†	
Parity	-1.82	4.35	-.05	-10.58	6.93	-0.42	.677	
Gestational age	-0.70	0.25	-.36	-1.20	-0.20	-2.84	.007**	
Infant sex	-6.25	4.12	-.19	-14.53	2.03	-1.52	.136	
HCC	5.26	10.09	.07	-15.02	25.54	0.52	.604	
SCC	-23.83	8.13	-.37	-40.18	-7.48	-2.93	.005**	
Time of day	5.05	4.24	.15	-3.47	13.56	1.19	.240	
EPDS score	-1.01	0.53	-.23	-2.08	0.06	-1.90	.064†	

Note: N = 57; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed), time of day is dummy coded as '0 = before 12 pm' and '1 = after 12 pm'; ** $p < .01$, † $p < .10$.

Table S 7

Results of the multiple regression models for infant protesting behavior *including the additional factors day of time and the EPDS score.*

	B	*SE*	β	95% CI lower	95% CI upper	*t*	*p*	R^2
Model 1								**.06**
Intercept	-191.00	194.85		-582.00	199.99	-0.98	.331	
Maternal age	0.67	1.38	.07	-2.09	3.43	0.49	.627	
Parity	8.26	11.32	.10	-14.46	30.98	0.73	.469	
Gestational age	0.72	0.65	.16	-0.59	2.03	1.10	.276	
Infant sex	14.49	10.86	.18	-7.31	36.28	1.33	.188	
Model 2								**.23**
Intercept	-331.18	204.28		-741.90	79.55	-1.62	.112	
Maternal age	1.13	1.36	.12	-1.61	3.87	0.83	.411	
Parity	-0.40	11.38	-.01	-23.27	22.48	-0.04	.972	
Gestational age	0.78	0.64	.17	-0.52	2.07	1.20	.235	
Infant sex	23.55	10.76	.30	1.92	45.18	2.19	.033*	
HCC	-27.43	26.35	-.14	-80.41	25.54	-1.04	.303	

SCC	55.24	21.25	.36	12.52	97.96	2.60	.012*
Time of day	-5.01	11.07	-.06	-27.26	17.25	-0.45	.653
EPDS score	1.37	1.39	.13	-1.43	4.18	0.99	.329

Note: N = 57; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed), time of day is dummy coded as '0 = before 12 pm' and '1 = after 12 pm'; * $p < .05$.

Table S 8

Results of the multiple regression models for maternal positive affect *including the additional factors day of time and the EPDS score.*

	B	*SE*	β	95% CI lower	upper	*t*	*p*	R^2
Model 1								**.08**
Intercept	60.84	50.91		-41.32	162.99	1.20	.237	
Maternal age	0.01	0.36	.00	-0.71	0.73	0.02	.983	
Parity	-3.49	2.96	-.17	-9.43	2.44	-1.18	.243	
Gestational age	0.12	0.17	.10	-0.22	0.47	0.72	.476	
Infant sex	-4.03	2.84	-.19	-9.72	1.66	-1.42	.162	
Model 2								**.10**
Intercept	51.26	58.14		-65.64	168.15	0.88	.382	
Maternal age	0.07	0.39	.03	-0.71	0.85	0.18	.857	
Parity	-3.31	3.24	-.16	-9.82	3.20	-1.02	.312	
Gestational age	0.11	0.18	.09	-0.25	0.48	0.62	.536	
Infant sex	-4.17	3.06	-.20	-10.33	1.98	-1.36	.179	
HCC	2.75	7.50	.05	-12.33	17.82	0.37	.716	
SCC	2.33	6.05	.06	-9.83	14.48	0.39	.702	
Time of day	1.69	3.15	.08	-4.65	8.02	0.54	.595	
EPDS score	0.36	0.40	.13	-0.44	1.15	0.90	.372	

Note: N = 57; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed), time of day is dummy coded as '0 = before 12 pm' and '1 = after 12 pm'; ** $p < .01$, * $p < .05$, † $p < .10$.

Table S 9

Results of the multiple regression models for match *including the additional factors day of time and the EPDS score.*

	B	*SE*	β	95% CI lower	upper	*t*	*p*	R^2
Model 1								**.13**
Intercept	248.34	100.36		46.96	449.72	2.48	.017*	

Maternal age	-0.83	0.71	-.16	-2.25	0.60	-1.17	.249	
Parity	-8.98	5.83	-.21	-20.68	2.73	-1.54	.130	
Gestational age	-0.68	0.34	-.27	-1.36	-0.01	-2.03	.048*	
Infant sex	-5.73	5.59	-.14	-16.96	5.49	-1.03	.310	
Model 2								**.34****
Intercept	313.71	100.47		111.70	515.71	3.12	.003**	
Maternal age	-1.03	0.67	-.20	-2.38	0.31	-1.55	.129	
Parity	-4.02	5.60	-.09	-15.27	7.23	-0.72	.476	
Gestational age	-0.65	0.32	-.26	-1.28	-0.01	-2.04	.046*	
Infant sex	-10.65	5.29	-.25	-21.29	-0.02	-2.01	.050*	
HCC	6.72	12.96	.07	-19.34	32.77	0.52	.607	
SCC	-31.55	10.45	-.38	-52.55	-10.54	-3.02	.004**	
Time of day	6.12	5.44	.14	-4.82	17.07	1.13	.266	
EPDS score	-1.20	0.69	-.22	-2.58	0.18	-1.75	.086†	

Note: N = 57; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed), time of day is dummy coded as '0 = before 12 pm' and '1 = after 12 pm'; ** $p < .01$, * $p < .05$, † $p < .10$.

Table S 10

Results of the multiple regression models for interactive reparation *including the additional factors day of time and the EPDS score.*

	B	*SE*	β	95% CI lower	95% CI upper	*t*	*p*	R^2
Model 1								**.13**
Intercept	-236.89	186.53		-621.07	147.28	-1.27	.216	
Maternal age	0.68	1.55	.08	-2.50	3.87	0.44	.662	
Parity	16.61	11.40	.27	-6.87	40.09	1.46	.157	
Gestational age	0.78	0.64	.23	-0.53	2.09	1.22	.233	
Infant sex	2.04	11.05	.04	-20.73	24.81	0.18	.855	
Model 2								**.39**
Intercept	-475.30	196.05		-883.02	-67.59	-2.42	.024*	
Maternal age	0.65	1.69	.08	-2.87	4.17	0.39	.704	
Parity	24.09	11.44	.40	0.31	47.88	2.11	.047*	
Gestational age	1,34	0.63	.40	0.02	2.66	2.11	.047*	
Infant sex	-4.00	13.66	-.07	-32.40	24.40	-0.29	.772	
HCC	19.50	37.91	.13	-59.34	98.34	0.51	.612	
SCC	23.46	23.89	.19	-26.23	73.14	0.98	.337	
Time of day	36.42	12.19	.59	11.07	61.77	2.99	.007**	
EPDS score	-0.26	1.40	-.04	-3.16	2.65	-0.19	.854	

Note: N = 30; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G;

log-transformed), time of day is dummy coded as '0 = before 12 pm' and '1 = after 12 pm'; ** $p < .01$, * $p < .05$.

Table S 11

Results of the multiple regression models for dyadic eye-contact *including the additional factors day of time and the EPDS score.*

	B	*SE*	β	95% CI lower	upper	*t*	*p*	R^2
Model 1								**.07**
Intercept	171.01	104.85		-39.38	381.41	1.63	.109	
Maternal age	-0.56	0.74	-.11	-2.04	0.93	-0.75	.456	
Parity	-7.83	6.09	-.18	-20.06	4.39	-1.29	.204	
Gestational age	-0.42	0.35	-.17	-1.13	0.29	-1.20	.238	
Infant sex	-4.44	5.84	-.10	-16.17	7.28	-0.76	.450	
Model 2								**.24**†
Intercept	212.61	109.56		-7.67	432.89	1.94	.058†	
Maternal age	-0.53	0.73	-.10	-2.00	0.94	-0.73	.471	
Parity	-3.83	6.10	-.09	-16.09	8.44	-0.63	.533	
Gestational age	-0.38	0.35	-.15	-1.08	0.31	-1.11	.272	
Infant sex	-8.45	5.77	-.20	-20.06	3.15	-1.47	.149	
HCC	-0.28	14.13	-.00	-28.69	28.14	-0.02	.984	
SCC	-25.35	11.39	-.31	-48.26	-2.44	-2.23	.031*	
Time of day	9.77	5.94	.23	-2.17	21.70	1.65	.106	
EPDS score	-0.61	0.75	-.11	-2.11	0.89	-0.82	.419	

Note: N = 57; parity is dummy coded as '0 = primiparous' and '1 = multiparous', sex is dummy coded as '0 = female' and '1 = male', HCC = maternal hair cortisol concentration (log-transformed), SCC = maternal salivary cortisol concentration indicated as area under the curve with respect to ground (AUC_G; log-transformed), time of day is dummy coded as '0 = before 12 pm' and '1 = after 12 pm'; * $p < .05$, † $p < .10$.