Developmental Review 57 (2020) 100922

Contents lists available at ScienceDirect

Developmental Review

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

Amniotic testosterone and psychological sex differences: A systematic review of the extreme male brain theory

Hui Xiong[a,*], Jordan B. Peterson[b], Stephen Scott[a]

[a] Institute of Psychiatry, Psychology and Neuroscience, Denmark Hill, King's College London, London, UK
[b] Psychology Department at the University of Toronto, Canada

ARTICLE INFO

Keywords:
Autism spectrum disorder
Sex differences
Amniocentesis
Fetal testosterone
Hormones
Extreme male brain

ABSTRACT

Background: Baron-Cohen (2002) proposed the Extreme Male Brain Theory (EMB) to suggest that foetal testosterone (FT) (1) is a component of the complex neurobiological aetiology of Autism Spectrum Disorder (ASD) and (2) accounts for its high male prevalence. The theory suggests that ASD is more common in males to an extreme manifestation of psychological maleness due to heightened testosterone exposure in the foetus.

Aim: To assess the EMB theory by reviewing cohort studies that directly assayed FT levels at 12–24 weeks of gestation in relation to subsequent ASD symptoms, ASD-related cognitions, social outcomes and playstyles prior to adolescence.

Method: A systematic term to subject heading search was conducted on Web of Science, Embase, PsycINFO, 'Ovid Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid Medline', PsycARTICLES Full Text, and ProQuest up to December 2019. Studies that included the extraction of foetal fluid and children of both sexes were assessed in compliance with STROBE guidelines. Additional articles were obtained by reference list screening.

Results: 22 FT-assay studies ($N = 2284$) containing EMB-associated traits as dependent variables were identified, including ASD symptoms, ASD-related cognition, sociality and playstyles. Their STROBE ratings ranged from 50% to 86.4%. FT significantly accounted for ASD-related traits beyond the child's sex in 3 of 4 studies. 4 out of 9 papers looking at sexed ASD-related cognitive-styles and 2 of 3 examining social outcomes showed significant FT effect. 2 of 6 found that FT accounted for significant variance in behavioral indices that differ on average between the sexes. Chi-square tests ($\chi^2(2, N = 22) = 4.46, P < .05$) demonstrated that researchers affiliated with Baron-Cohen are significantly more likely to generate results fully supportive of EMB, with 25% ($N = 3, P < .05$) of positive findings produced by independent authors. Homogeneity of data did not account for this.

Conclusion: The certainty with which FT was established as an agent in sexual differentiation varies by the psychological variable in question, but none of the conclusions were supported by an adequate number of studies. Nevertheless, this review yields the following preliminary conclusions, which can be tested in future research. FT plays a plausible role in driving social and non-social ASD-related cognition as well as ASD symptoms across the sexes. FT accounts for gender differences on eye contact frequency and value-laden proposition use and mediates the narrowing of interest toward systems and exerts sex-specific effects on numerical and language abilities, though these studies require independent replication. The role of FT on the differentiation of play is consistently non-significant. Where an effect exists, it is largely dwarfed by the effect of sex and hence it is equivocal that second trimester FT affects play. Biological

* Corresponding author.
E-mail address: h.xiong.16@ucl.ac.uk (H. Xiong).

https://doi.org/10.1016/j.dr.2020.100922
Received 25 June 2019; Received in revised form 11 June 2020
Available online 02 July 2020

implications for sex differences are considered and more lifespan longitudinal amniocentesis studies are suggested to pursue greater clarity in the empirical bases of EMB.

Introduction

Background

Autism Spectrum Disorder (ASD) is defined by the fifth Diagnostic and Statistical Manual of Mental Disorders (DSM) as social-communicative deficits, sensory reactivity, and restricted/ repetitive patterns of interests (APA, 2013; Tavassoli et al., 2017). Baron-Cohen (2002) formalized the Extreme Male Brain theory (EMB) to account for the profile of ASD in relation to broader sex differences. As a neurodevelopmental condition accompanied by areas of cognitive difficulties and strengths, the autism phenotype, according to Baron-Cohen (2002), is one of strengths in sex dimorphic areas that are heightened in males and weaknesses on facets in which females are advantageous. The EMB asserts that ASD and its related traits are the result of upregulated foetal testosterone (FT) concentrations in the amniotic fluid, not mutually-exclusive of other neurobiological, genetic and social factors (Stodgell, Ingram, & Hyman, 2000). This review will aim to systematically assess these premises of EMB – 1) in trait sex differences; and 2) possible aetiology rooted in FT– according to the current state of the literature.

The psychological dimensions that show sex disparity, as suggested by the EMB, can be parsed into three categories: cognitive, social functioning, and behavioural. These domains will be outlined here in turn. Within EMB, systemizing is a major intellectual feature underlying ASD. It is a construct conceived by Baron-Cohen (2002) to be the competency and motivating drive underlying the effort to understand input-operation-output patterns. It is a prenatally primed propensity to attend to pivotal elements of a system and observing the outcome when these components are varied or manipulated. Doing so enables the replication of if-then regularities of key factors within a given construct – be they technology, plants or otherwise – that affects the overall functioning (Baron-Cohen, 2008). It is heightened in neurotypical males compared to females in some but not all samples (Auyeung, Taylor, Hackett, & Baron-Cohen, 2010; Caldwell-Harris & Jordan, 2014; Xiong, Swift & Peterson, 2016; Yang & Barth, 2015). Those with high-functioning ASD display this thinking-style to a yet higher extent than neurotypically developing men, suggesting an exaggeration of the male profile (Baron-Cohen et al., 2003). Social cognition and person-orientation, while superior in females in non-clinical populations, are impaired among those with ASD (Baron-Cohen et al., 2014). Rough-and-tumble (RAT) play is also more common among boys than girls, and it is one of the few social activities pursued by those diagnosed with ASD according to Wing (1985). These emergent profiles of ASD as the heightened phenotypes of the neurotypical male psychology provide preliminary validation for EMB.

Baron-Cohen (2002) identification of FT as the explanatory agent derives from the precedent of Geschwind and Galaburda (1985). The latter proposed a theory of androgen-induced brain masculinization wherein FT delayed left-hemisphere development and impaired immune functioning (McManus & Bryden, 1991). Testosterone levels differ most dramatically between the sexes during the eighth to twenty-fourth week periods of gestation, a period anticipated to be accompanied by heightened sexual differentiation (Auyeung et al., 2010). While various proxies of FT-mediated masculinization have surfaced (e.g. umbilical cord blood assays), providing different readings on hormone levels, amniocentesis is the measure of choice that most faithfully assays testosterone levels in the womb at this developmental juncture (Lutchmaya, Baron-Cohen, Raggatt, Knickmeyer, & Manning, 2004; Rodeck, Gill, Rosenberg, & Collins, 1985). It is routinely conducted by obstetricians from 12 weeks in gestation and not done prior to this period. The procedure carries a 1% increased risk of miscarriage and conducting it at earlier gestational periods introduce risks for further complications (Akolekar, Beta, Picciarelli, Ogilvie, & D'antonio, F. , 2015). Prospectively linking FT levels to later facets of child psychology will be key to appraising whether FT-induced physiological sexual differentiation at this stage is a significant factor in driving psychological sex differences. At the extreme of maleness, according to EMB, an infant is at greater risk for ASD.

FT-induced masculinization offers a parsimonious account for the gender-prevalence of a range of mental disorders. ASD is among a few developmental conditions that have a male preponderance. While the incidence of boys receiving a diagnosis in ASD and Asperger Syndrome is fourfold and elevenfold relative to girls, respectively (Chakrabarti & Fombonne, 2001; Gillberg, Cederlund, Lamberg, & Zeijlon, 2006), other clinical diagnoses associated with early onset and chronic trajectories are also more common among boys. These include obsessive compulsive disorder subtypes, attention deficit hyperactivity disorder, and persistent antisocial behaviour (Rutter, Caspi & Moffitt, 2003). Like subjects with ASD, aggression or social isolation – traits that are cross-culturally higher in males relative to females – are especially heightened in these clinical populations (Jensen, Martin, & Cantwell, 1997; Mathis et al., 2011; Weisberg, DeYoung & Hirsh, 2011). The finding that those with ASD also manifest an extreme in traits that are elevated in males relative to females implicates a sex-specific vulnerability. Testing the EMB would help to make the first steps into extricating shared psychological risk factors across disorders that show sex-specific impairments.

Scope

An assessment of FT levels in relation to sex differences across studies in children's ASD symptoms, traits related to ASD and that which extend into the broader sex dimorphic phenotype will make clear the degree to which empirical data accord with EMB's predictions for each domain. These general categories are 1) ASD symptoms; 2) cognitive styles; 3) empathy and theory of mind social functioning indices; and 4) gendered behavioural styles.

Various outcome measures are included by their association with ASD. Researchers have made specific empirical or theoretical associations of constructs in contributing to ASD functions relevant to one of the above four domains. Restricted interests is subsumed under intellectual functioning as it is asserted to underlie systemizing, which is central to EMB's characterization of ASD. Knickmeyer et al. (2005) suggest that the former is the capacity for prolonged examination on detailed aspects of a phenomenon – a necessary skill to extrapolate the laws of a given system. Cerebral lateralization, as indexed by handedness and dichotic listening, pertains to language functioning, visuospatial skills and other cognitive domains that are impacted in ASD (Finegan, Niccols, & Sitarenios, 1992; Grimshaw, Bryden, & Finegan, 1995). Autistic savants, for example, demonstrate highly enhanced perceptual processing skills, which are matched by profound linguistic deficits (Mottron, Dawson, Soulières, Hubert, & Burack, 2006). RAT is included under ASD-related behaviours because it is heightened in males, across species, and is described by Wing (1985) as the only preferred social activity among autists (Hines, 2006). Likewise, delay of gratification and aggressiveness are also considered due to the comorbidity of ASD with Attention Deficit Hyperactivity and Conduct Disorder (Craig et al., 2016; Stevens, Peng, & Barnard-Brak, 2016).

Where there is a conflict in classification, the case is filed into the more exclusive category. An example is the Reading the Mind in the Eyes task, which is a performance measure of theory of mind wherein subjects extrapolate the emotions of another by a picture of the eyes. It displays a stronger correlation with verbal fluency than cognitive empathy and can be classed into either cognition or social functioning (Baker, Peterson, Pulos, & Kirkland, 2014). However, not all thinking-styles pertain to social functioning even though some relationship measures also pertain to information processing. Thus, its distinguishing feature is that it is a measure of social deficits.

The EMB has constellated a sufficiently substantial body of research in these areas for its formal appraisal. Several narrative reviews have been devoted to the subject (e.g. Baron-Cohen, 2005) and scholars have not been consistently supportive of EMB. With an admixture of findings raised from several sources, a systematic review will allow the signal, where it exists, to come to the fore from the noise (Kung et al., 2016; Whitehouse, 2016). Since the theory's conception by Baron-Cohen (2002), the majority of EMB's empirical testing by FT data have come from members of his research network. This potential source of bias will also be examined. A screening of databases and the International Prospective Register of Systematic Review (PROSPERO) indicates that no systematic review has been completed to assesses the predictions of EMB.

The current review will be based on children's ASD traits in a focused manner, limiting the scope in several ways: 1) As no studies to date have investigated the psychological contingency of high FT in post-pubertal adolescents and adults, the studies reviewed will be narrowed to children. 2) Moreover, since the body of research of FT and psychological outcomes are small, many studies focusing on ASD traits and related behaviour (e.g. language lateralization) have not received replication. A systematic review of the small but emerging research available, rather than a meta-analysis on a higher number of studies for a specific outcome variable, is the most suitable approach to assess the state of the literature. 3) The causal role of FT in facilitating psychological and behavioural sexual differentiation in animal models have been the subject of previous reviews (e.g. Hines, 2006). Inferring their relevance to autists, however, requires a systematic analysis confined to human samples.

The review will also exclude 2nd to 4th digit ratios and perinatal testosterone levels in umbilical cord blood as measures for FT. These finger lengths are a dubious index of FT. Since FT levels are not stable across gestation, the differentiation of the digits is not specifically associated with androgen exposure during the second trimester critical period but begins prior to this time. The index is not necessarily anticipated by the EMB theory to predict ASD-related cognitions (Voraceck & Dressler, 2006). Moreover, other hormone measures such as prenatal maternal and perinatal umbilical cord serum have not been demonstrated to reflect second-trimester FT levels and cannot be assumed to related due to hormonal variation across gestation (Van De Beek, Thijssen, Cohen-Kettenis, Van Goozen, & Buitelaar, 2004). Introducing these measures into the review would deviate from accurately capturing FT levels during the 8th to 14th weeks of gestation as would be possible with amniocentesis.

Congenital medical conditions that manifest in raised testosterone levels in utero were also not within the scope of the current review. Abnormal FT has been, for example, observed in such disorders as congenital adrenal hyperplasia (CAH), and correspond to increased rates of rough-and-tumble play and male-typical toy preferences (Hines, 2006). While these studies would apparently contribute to the current assessment of EMB, their inclusion would obstruct a clear illustration of the effect of FT on the brain. It is unclear using these samples whether the FT levels have been abnormally elevated only within the period of sexual dimorphism, within periods outside of the 2nd trimester and/or reflect greater baseline levels throughout gestation. CAH is also associated with abnormal maternal response to the offspring, adding a complex layer of psychosocial confounding (Slijper, 1984). The association between FT and sex-typed behaviours obtained within subjects with endocrine disorders does not clearly contribute to understanding the general aetiology of ASD. Their relevance to the validation of the EMB theory is therefore limited.

For a psychological domain to be in strict accordance with the EMB, two criteria must be met: 1) it must significantly differ between neuro-typical males and females; and 2) it can be significantly modelled in regression analyses by FT levels independently of gender. That is to say that for a trait to be considered sex-dependent, it must first be demonstrated that the degree that the phenotype manifested is correlated with gender. This shows that the trait is sex-linked. If FT taken during gestation then anticipates the future manifestation of the psychological facet, controlling for the child's sex, then EMB is supported, demonstrating that FT partially or fully mediates this dimension of sex differences. The finding that FT associates with psychological measures when no sex differences are detected does not argue against EMB, per se, as FT may interact with existent neurophysiological architecture to facilitate gender-dependent differentiation. FT, in this case, acts in a manner contingent upon the child's gender and offers partial assent to the theory. If the effect is non-existent or disappears upon controlling for sex, then the study would not support the role of FT. This finding would raise the possibility that FT does not impact the relevant sex-typical dimensions or is merely a covariate of sex and a statistical artefact, respectively (Fine, 2005).

Table 1
Characteristics of included studies.

Main Outcome	Significant Outcome Measure Sex Differences (Cohen's D)	FT Correlations with Main Outcome and Interactions	Controlled Confounds	Sample Characteristics: Participation Rate, Age, Sample Size, Respondents	Author (Year) Country	Description for Participant Source	Support for EMB	STROBE Rating
Foetal testosterone studies and sex disparities in ASD symptoms								
The Childhood Autism Spectrum Test (CAST) and the Child Autism Spectrum Quotient (AQ-Child)	Male > Female on AQ-Child (d = .87) and CAST (d= 0.30)	AQ-Child total scores were modeled by FT levels (β = .27, P < . 01), child sex (β = .21, P < . 01) and presence of older sister (β = .13, P < . 05). Within sex analyses revealed the significant contribution of FT was maintained after confounds are controlled. Social skills, mind reading, attention to detail and imagination were correlated with FT level and sex FT was the only significant predictor in modeling CAST. Within sex regressions indicated that FT was related to CAST scores in boys but not in girls.	Gestational age at amniocentesis, maternal age, parental education level, presence of older siblings, child's age	51.9%, 6–10 years, n = 235, Mothers Middle Childhood	Auyeung, Baron-Cohen, Ashwin, Knickmeyer, Taylor & Hackett 2009 England	Mothers who had undergone amniocentesis in Cambridge between 1996 and 2001	Positive	81.8%
The Quantitative Checklist for Autism in Toddlers (Q-CHAT)	Boys > Girls on Q-CHAT (d = .46)	FT was the only variable retained in the regression model for Q-CHAT (β = .49, P < . 001). FT was retained as significant predictors in both boys and girls	Gestational age at amniocentesis, maternal age, parental education, presence of older siblings, child's age	45.6%, 18 to 24 months, n = 129, Parent Infancy	Auyeung et al., 2010 England	700 mothers in Cambridge from January 2004 and July 2006	Positive*	77.3%
Q-CHAT	Boys > Girls on Q-CHAT (d = .92)	Multistep hierarchical regression retained FT (β = .49, P < . 001) as the only significant predictor for Q-CHAT. FT correlates significantly with Q-CHAT in boys (rho = .51, P < . 05) and girls (rho = .57, P < . 05)	Sex, postnatal Testosterone, Gestational Age, Child Age, Maternal Age, Parent Education, Older Sister, Older Brother	45.2%, 18–35 months, n = 35 Infancy	Auyeung et al., 2012 England	700 mothers in Cambridge from January 2004 and July 2006	Positive*	54.6%
CAST	**Boys > Girls in CAST (d = .44)**	**FT does not correlate with CAST in boys, girls or the two genders combined.**	**Vocabulary Subtest of Wechsler Intelligence Scale for Children (WISC-IV) and Age**	**86%, 3–5 years, n = 92, Parent Middle Childhood**	Kung et al., 2016 **England**	**Mothers who had undergone amniocentesis at Imperial College, NHS trust**	**Null**	**63.6%**
Foetal Testosterone studies and Sex Differences in ASD-related Cognition								
The Revised Systemizing Quotient for Children (SQ-C)	Males > Females on SQ-C (d = .69)	Stepwise linear regression for SQ-C resulted in FT reaching significance atβ = .37, t(172) = 5.29, P < . 01	Gestational age at amniocentesis, parental age, level of parental education, number of siblings	46%, 6–9 years, n = 204, Parent Middle Childhood	Auyeung et al., 2006 England	As part of the Baron-Cohen, Lutchmaya, & Knickmeyer (2004) long-term study, participants were mothers who had undergone amniocentesis in Cambridge	Positive*	77.3%
							Partial	54.5%

(*continued on next page*)

Table 1 (*continued*)

Foetal testosterone studies and sex disparities in ASD symptoms

Main Outcome	Significant Outcome Measure Sex Differences (Cohen's D)	FT Correlations with Main Outcome and Interactions	Controlled Confounds	Sample Characteristics: Participation Rate, Age, Sample Size, Respondents	Author (Year) Country	Description for Participant Source	Support for EMB	STROBE Rating
Children's Embedded Figures Task (EFT), ball targeting task, and computerized mental rotation task	Males > Females on Mental Rotation and EFT (d = 1.15)	Hierarchical regression for mental rotation retained child sex and age but not FT Hierarchical multiple regression for EFT retained FT atβ = .63, P < . 001. FT remained a significant predictor within each sex.	Parent education, presence of older siblings, gestational age, number of siblings	Unknown, 7–10 years, n = 64, Child Middle Childhood	Auyeung et al., 2012 England	Mothers who had undergone routine amniocentesis in the Cambridge region from 1996 to 1999		
Language expression; Language comprehension; problem solving, conceptual grouping, visual-motor integration; embedded figures; counting and sorting; number questions; verbal memory; General Cognitive Index	**None**	**For girls, FT displayed a quadratic inverted-U relationship with Language Comprehension (P = .003) and Conceptual Grouping scores in regression models (P = .019). Higher FT linearly associated with lower scores on Counting and Sorting (P = .007) and Number Questions (P = .03)** **In boys, FT did not predict Language Comprehension, Conceptual Grouping, Visual-Motor Integration, Counting and Sorting or Number Question scores**	**Parental social class, maternal IQ, mothers' year of education, and obstetric complications**	**38.9%. 4 years, n = 60, Child Middle Childhood**	Finegan et al., 1992 **Canada**	**Prospective mothers at the Antenatal Genetics Clinic at the Toronto General Hospital, in continuation of** Finegan et al. (1984)	**Partial**	**86.4%**
Handedness, The Fused Dichotic Words Test (FDWT), Emotional Word Test (EWT)	**None**	**FT accounted for 17% of handedness variance and 19% FDWT laterality in girls. Higher FT associated with greater right-handedness and left-hemisphere language representation. FT responsible for 23% of laterality for EWT in boys. FT correlated with greater right hemisphere advantage in EWT processing.**	**Gestational age, birth stress**	**Study 1: Unknown, 10 years, n = 53, Child and Parents Study 2: Unknown, 61.4%, 10 years, n = 43, Child and Parents Middle Childhood**	Grimshaw et al., 1995 **Canada**	**Prospective mothers at the Antenatal Genetics Clinic at the Toronto General Hospital, in continuation of** Finegan et al. (1984)	**Partial**	**63.6%**
Communicative development inventory (CDI)	Girls > Boys on vocabulary size at 18 (d = .066) and 24 months (d = .63)	FT was a significant predictor of vocabulary size across the sexes at 18 and 24 months	Estradiol, alpha-foetoprotein, sex, number of siblings, maternal age, paternal age, parental education	Unknown, 18 – 24 months, n = 87, Child and Parent Infancy	Lutchmaya et al., 2001 England	Mothers who had undergone routine amniocentesis in the Cambridge region from June 1996 and June 1997	Positive*	63.6%
Children's Communication Checklist (CCC)	Male and Females did not differ on relationship quality Males > Females on restricted interest (RI) (d = .69)	FT was the only significant variable in negatively predicting social relationship quality The variance of RI that FT accounted for was beyond the child's sex. Higher FT was associated with greater RI.	Gestational age at amniocentesis, fetal estrogen, prenatal alpha-fetoprotein, maternal age, maternal education, paternal age, number of older siblings	Unknown, 4.0 – 4.25 years, n = 58, Child and Experimenter Early Childhood	Knickmeyer et al., 2005 England	Mothers who had undergone routine amniocentesis in the Cambridge region from June 1996 and June 1997	Positive*	63.6%

(*continued on next page*)

Table 1 (*continued*)

Foetal testosterone studies and sex disparities in ASD symptoms								
Main Outcome	Significant Outcome Measure Sex Differences (Cohen's D)	FT Correlations with Main Outcome and Interactions	Controlled Confounds	Sample Characteristics: Participation Rate, Age, Sample Size, Respondents	Author (Year) Country	Description for Participant Source	Support for EMB	STROBE Rating
Dichotic listening	**Right ear advantage over the left in boys $d = 1.86$) but no difference in girls**	**FT predicted stronger left hemisphere language lateralization across the sexes FT positively related percent correct in right-focused condition in girls, and negatively to left-focused condition in boys**	**Hand preference**	**Unknown, 6 years, n = 54, Child** Middle Childhood	Lust et al., 2010, **The Netherlands**	**A subset of the** Van De Beek et al. (2004) **longitudinal study**	**Positive***	**50%**
Handedness, language lateralization	**Unreported**	**FT predicted decreased strength of handedness ($p < .01$)**	**IQ (Verbal, Situations and Analogies)**	**Unknown, 6 – 7.5 years, n = 65, Child and Experimenter** Middle Childhood	Lust et al., 2011, **The Netherlands**	**Children followed up as part of the** Van De Beek et al. (2004) **longitudinal study**	**Negative**	**50%**
Mental Rotation, Chimeric Faces, Word Generation Task, Pubertal Testosterone	**Males > Females on Mental Rotation ($d = 0.53$)**	**On the Chimeric Faces and Mental Rotation tasks, FT moderated the association of pubertal testosterone with hemispheric lateralization in boys. Among the low FT, pubertal testosterone positively related to right hemisphere specialization whereas the reverse pattern was the case among those with high FT.**	**Pubertal stage, sex of child**	**Unknown, 15 years, n = 60, Child and Experimenter Adolescent**	Beking et al., 2018 **The Netherlands**	**Children followed up as part of the** Van De Beek et al. (2004) **longitudinal study**	**Positive***	**55%**
Foetal Testosterone studies and Sex Differences in Social Functioning								
Study 1: The Empathizing Quotient (EQ-C) Study 2: Reading the Mind in the Eyes (Eyes-C)	Study 1: Females > Males on EQ, $d = .76$ Study 2: Eyes-C did not significantly differ between boys and girls	Study 1: FT was significant when regressing on EQ only within boys ($r(99) = -.35$, $P < .01$) Study 2: FT accounted for unique variance ($\beta = -.68$, $P = .001$) when included in hierarchical regression, controlling for child's age. FT significantly correlated with Eyes-C in boys ($r(38) = -.42$, $P < .01$) and girls ($r(34) = -.29$, $P < .05$)	Sex of child, gestational age at amniocentesis, maternal age, paternal age, number of siblings, maternal education and child's age	Study 1: Unknown, 6 – 9 years, n = 193, Parent Study 2: Unknown, 6 – 9 years, n = 78, Child Middle Childhood	Chapman et al., 2006 England	Mothers who went through amniocentesis in Cambridgeshire, Norfolk or Suffolk between June 1996 and June 1999, as part of the Cambridge longitudinal study	Partial	81.8%
Mental state and Affective state terms in children's narrative to animation	Females > Males on number of affective state terms ($d = .82$) Males > Females in terms of neutral proposition production($d = .63$)	FT was the only predictor retained in the regression on intentional propositions. FT was inversely associated with intentional propositions. FT explained the variance of neutral proposition beyond sex. Higher FT was related to greater use of neutral propositions.	Prenatal estrogen, fetal alpha-fetoprotein, sex, gestational age at amniocentesis, maternal age and number of siblings	Unknown 4.0 – 4.25 years, n = 39, Child and experimenter Early Childhood	Knickmeyer et al., 2006 England	Children whose mothers took part in long-term FT study	Positive*	68.2%
							Positive	77.3%

(*continued on next page*)

Table 1 (*continued*)

Foetal testosterone studies and sex disparities in ASD symptoms								
Main Outcome	Significant Outcome Measure Sex Differences (Cohen's D)	FT Correlations with Main Outcome and Interactions	Controlled Confounds	Sample Characteristics: Participation Rate, Age, Sample Size, Respondents	Author (Year) Country	Description for Participant Source	Support for EMB	STROBE Rating
Eye contact frequency between parent and child	Female > Male in eye-contact frequency (d = .53)	FT ($B = -2.1$, $P < .001$) and FT squared was retained in regression. FT displayed an inverse-U relationship with eye-contact frequency	Estradiol, alpha-fetoprotein, gestational age at amniocentesis, number of siblings, maternal age at birth, paternal age at birth, level of education attained by parents	Unknown 12 months, n = 70, Child and experimenter Infancy	Lutchmaya et al., 2002 England	Mothers who had undergone routine amniocentesis in the Cambridge region from June 1996 and June 1997		
Foetal Testosterone studies and Sexual Disparities in Play and Behavior								
Preschool Activities Inventory (PSAI)	Males > Females on PSAI (d = 2.92)	Hierarchical regression for PSAI was significant at$F(2, 181) = 250.11$, $P < .001$, with sex ($\beta = .77$, $P < .001$) and FT ($\beta = .14$, $P < .01$) being significant predictors.	Gestational age at amniocentesis, maternal age, maternal education, child's age, sex, sex and fetal testosterone interaction	46.0%, 6.38 – 10.30 years, n = 212, Mothers Middle Childhood	Auyeung, Baron-Cohen, Auyeung et al., 2009b England	As part of the Baron-Cohen, Lutchmaya, & Knickmeyer (2004) long-term study, participants were mothers who had undergone amniocentesis in Cambridge	Positive*	68.2%
Mental Rotation and Spatial Play Experience	**Boys > Girls in Frequency of Spatial Play (Cohen's d is unknown)**	**FT positively predicted rotation rate ($r(9) = .67$, $P = .02$) and negatively with response time ($r(10) = -.63$, $P = .03$) among the rotational-strategy female subgroup Spatial play was not related to FT in boys**	**None**	**Unknown, 7 years, n = 60, Child Middle Childhood**	Grimshaw et al., 1995 **Canada**	**Children followed up as part of the** Van De Beek et al. (2004) **longitudinal study**	**Partial**	**63.6%**
Child Game Participation Questionnaire (CGPQ)	Boys > Girls on Masculinity scale, d = 3.6. Girls > Boys on Femininity scale,d = 3.2	FT did not account for variance on Masculinity scale beyond sex. Sex and Sex × FT interaction was significant in predicting Femininity scores. The FT-Femininity correlation approached significance in girls, $r(21) = .36$, $P = .10$.	Fetal estrogen, prenatal alpha-fetoprotein, sex, gestational age at amniocentesis, maternal age, maternal education, gender of older siblings	33.1%, 4.75 – 5.8 years, n = 53, Mothers Early Childhood	Knickmeyer, Wheelright et al., 2005 England	Mothers who had undergone routine amniocentesis in the Cambridge region from June 1996 and June 1997	Null	72.7%
Structured Play Observation	**Boys > Girls in percent of time spent on Masculine toys, d = .53. Girls > Boys in percent of time spent on feminine toys,d = .35**	**FT did not predict Masculine or Feminine play. Progesterone, number of older brothers and parental educational level significantly modeled Masculine play. Progesterone levels did not significantly differ between the sexes**	**Maternal Serum Sex Hormone, Bayley Scales of Infant Development II (BSID II), age of mother at amniocentesis, parental education, number of older brothers and sisters**	**88.2%, 13 months, n = 126, Child, mothers, and experimenter Infancy**	van de Beek et al., 2009 **The Netherlands**	**Unknown**	**Null**	**86.4%**
Delay of Gratification (DoG task)	**Girls > Boys in DoG**	**Boys showed higher FT than girls FT predicted worse delay of**	**Preschool and Kindergarten Behavior Scales-II (PKBS-II)**	**Unknown, 40 months,**	Körner et al., 2019 **Germany**	**Mothers who underwent amniocentesis**	**Positive**	54.6%

(*continued on next page*)

Table 1 (*continued*)

Foetal testosterone studies and sex disparities in ASD symptoms								
Main Outcome	Significant Outcome Measure Sex Differences (Cohen's D)	FT Correlations with Main Outcome and Interactions	Controlled Confounds	Sample Characteristics: Participation Rate, Age, Sample Size, Respondents	Author (Year) Country	Description for Participant Source	Support for EMB	STROBE Rating
		gratification, $r(79) = -.26$, $P < .04$. Analysis was not possible for girls due to the majority FT levels that not reaching detection therein		**n = 122, Children Early Childhood**		**between 2010 and 2012 in Dusseldorf, German**		
Interests, Activities, & Temperament Questionnaire – II (IATQ-II)	**Boys > Girls on levels of Aggression (d = .41) and Activity (d = .50)**	**FT did not predict either factors in boys or girls**	**Age at testing, Weschler Intelligence Scale – Vocabulary, Wechsler Preschool and Primary Scale of Intelligence**	**Unknown, 3–5 years, n = 92, Parent-Reported Early Childhood**	Spencer et al., 2017 **England**	**Mothers who had undergone amniocentesis at Imperial College, NHS trust**	**Null**	72.3%

Study data not affiliated with Simon Baron-Cohen are highlighted in bold. * indicate that the Positive results derived from samples that have supported EMB in a different study.

Method

Guidelines were adopted from the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA), detailed in Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group (2009), which contains a 27-item checklist of reporting or assessment criteria in each section of the review paper.

Inclusion and exclusion criteria

Conditions required to be fulfilled to be included in the review were: 1) the sample must consist of children from both sexes; and 2) FT is extracted through amniocentesis. All other studies were excluded.

Search strategy

The database searches were conducted three times, first in December 2017, again, in compliance to editors' requests, in December 2019 and a third time in May 2020 on Web of Science. The first two were on the OvidSP bibliographic databases, including Embase; PsycINFO; Ovid Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) ALL; and PsycARTICLES Full Text. This method covered the existing literature therein from 1946 up to May 2020.

The Population, Intervention, Comparison, and Outcome (PICO) of targeted studies were neuro-typical children, foetal testosterone, male versus female, and autism-related individual differences, respectively. To facilitate as inclusive a research under such criteria, the Medical Subject Heading (MESH) thesaurus was consulted to generate multiple search terms from the same concepts. Due to the overly large number of articles that results from searching with foetal testosterone-related lexicon, such words were confined to mapping only onto article titles. The keywords used were: (Adolescent or child or children or youth or infant* or neonate*) AND (Foetal testosterone or prenatal testosterone or gestational testosterone or uterine testosterone or amniotic testosterone or foetal androgen or prenatal androgen or gestational androgen or uterine androgen or amniotic androgen or foetal hormone or prenatal hormone or gestational hormone or uterine hormone or amniotic hormone or amniocentesis or extreme male brain or prenatal sex hormones) [title only] AND (Boy* or girl* or sex difference$ or gender difference$ or sex differentiation or gender differentiation male* or female*) AND (Social skills or theory of mind or mentalizing or empath* or care or visuospatial ability or perspective taking or social cognition or cognitive skill or cognitive ability social relationship communication or social perception or vocabulary size or eye gaze or eye contact or systemizing or autistic trait* or cerebral lateralization or language lateralization or behavior or behaviour or play) (see Appendix A). These terms were then applied to the ProQuest grey literature.

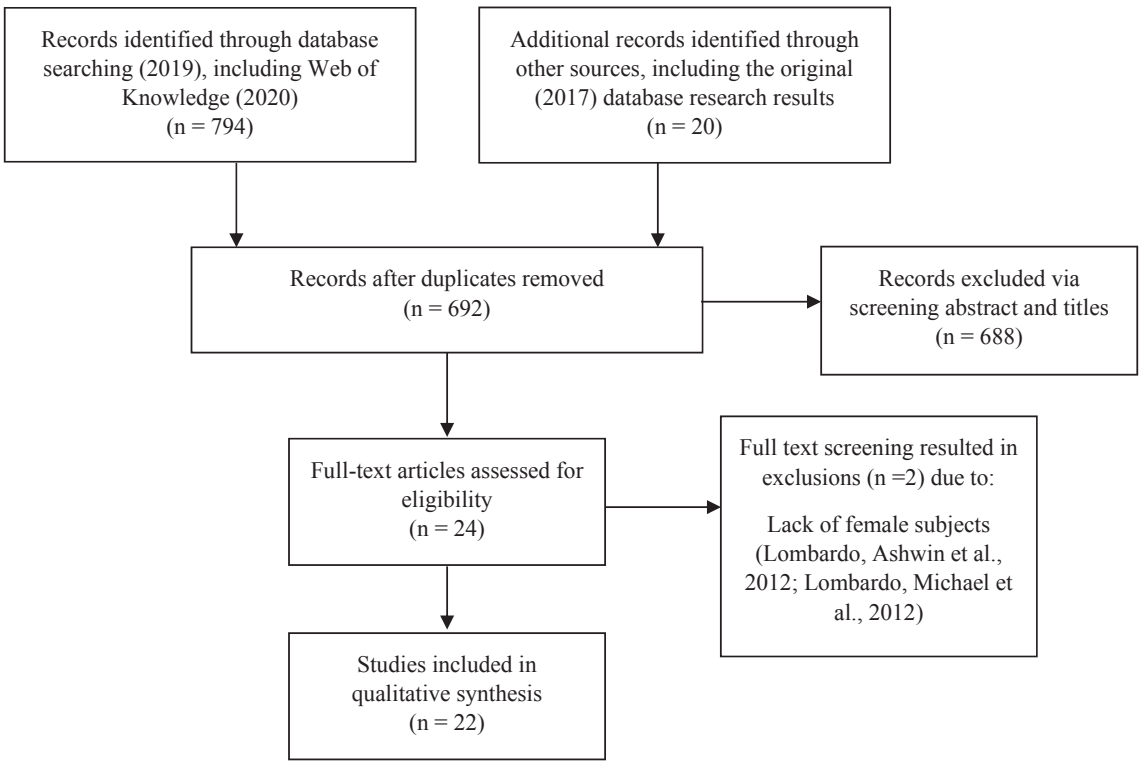

Fig. 1. Study flow.

Data extraction

The primary characteristics of studies that met inclusion criteria were compiled by the main researcher (see Table 1). Participation rate was the quotient of the sample presented divided by the total number of women receiving amniocentesis. Age of children, number of participants, and respondent identity (i.e. whether it was observer report or performance of the child) were recorded. Control variables, main outcome measure, and whether the latter displayed sex differences was documented. Results from regressing FT and confounds onto the main outcome were reported if FT emerged as a significant predictor. Each study was then judged by the first author on the extent to which it supported EMB and were graded Positive, Partial, Null and Negative in descending degree of concordance.

Risk and bias assessment

The risk of bias for individual articles was assessed by the primary author using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (Vandenbroucke et al., 2007). The 22-item checklist for cohort studies were selected. Consistent cut-offs were applied to systematically determine whether a condition was met. On the one hand, for criteria that contain multiple sub-facets, violation of even one would result in a deduction. For example, the statistical methods section in the STROBE contains five features. If one of them was missing in the paper, it would have been downgraded in this domain. On the other hand, any condition that was fulfilled once in the paper, however briefly, was given a point. Overall ratings were computed based on the quotient between the number of criteria met and 22.

Researcher bias was tested using Pearson's chi-square test of independence to examine the relation between the source of the dataset – i.e. whether it were from members of the Baron-Cohen group or not – and the degree to which the results were supportive of the EMB hypothesis – i.e. null, negative or partial and positive. The sources of the samples were reported in Table 1. Positive findings that were found multiple times in the same subject pool were flagged, tabulated and reported below.

Results

Study selection

The searches produced a total of 792 articles (see Fig. 1 for the flow of study selection). Searches conducted in 2017, 2019 and 2020 were compiled into the screening. Thereafter, 120 studies were removed due to duplication and 668 to screening. In the remaining 24, 2 were excluded due to single-sex samples, finalizing with 22 works included in this review ($N = 2284$).

Study characteristics

Table 1 displays the study characteristics. 4 reported ASD symptom scores (Auyeung et al., 2010, 2012; Auyeung, Baron-Cohen, Ashwin, Knickmeyer, Taylor & Hackett, 2009; Kung et al., 2016), 9 ASD-related cognition (Beking et al., 2018; Finegan et al., 1992; Grimshaw et al., 1995; Lutchmaya, Baron-Cohen, & Raggatt, 2001; Knickmeyer et al., 2005; Auyeung et al., 2006; Lust et al., 2010, 2011; Auyeung et al., 2012), 4 focused on social outcomes (Auyeung, 2006; Chapman et al., 2006; Knickmeyer, Baron-Cohen, Raggatt, Taylor, & Hackett, 2006; Lutchmaya, Baron-Cohen, & Raggatt, 2002), and 6 on behavioural and play-styles (Grimshaw, Sitarenios, & Finegan, 1995; Knickmeyer, Wheelright et al., 2005; Auyeung et al., 2009; Körner et al., 2019; Spencer et al., 2017; van de Beek, van Goozen, Buitelaar, & Cohen-Kettenis, 2009). 12 of these studies were conducted by Baron-Cohen's group (Lutchmaya et al., 2001, 2002; Knickmeyer et al., 2005; Knickmeyer, Wheelright et al., 2005; Auyeung, 2006; Auyeung et al., 2009a, 2009b, 2010, 2012; Chapman et al., 2006; Knickmeyer et al., 2006). 13 were conducted in England (Lutchmaya et al., 2001a, 2002b, 2002c; Knickmeyer et al., 2005; Knickmeyer, Wheelright et al., 2005; Auyeung 2006; Auyeung et al., 2009a, 2009b, 2010, 2012; Chapman et al., 2006; Knickmeyer et al., 2006; Kung et al., 2016; Spencer et al., 2017), 1 Germany (Körner et al., 2019), 3 in Canada (Grimshaw et al., 1995a, 1995b; Finegan et al., 1992), and 4 in the Netherlands (Beking et al., 2018; van de Beek et al., 2009; Lust et al., 2010, 2011).

The effect of FT on sex differences in ASD symptoms

Among the four articles that reported ASD symptoms in neurotypical children, all of them detected greater severity of ASD in males relative to females (Auyeung et al., 2009a, 2010; Kung et al., 2016). FT predicted ASD symptoms beyond sex in three studies (Auyeung et al., 2009a, 2010, 2012). FT significantly modelled Autism Quotient (AQ) scores within each sex in Auyeung et al. (2009a), though this was not the case with the Childhood Autism Spectrum Test (CAST). FT was only associated with CAST scores in boys. Auyeung et al. (2010) and Auyeung et al. (2012) found that males were higher on the Quantitative Checklist for Autism Test (Q-CHAT) than females. FT was the sole predictor of Q-CHAT in regression. Auyeung et al. (2009a), Auyeung et al. (2012) and Auyeung et al. (2010) are concordant with FT masculinizing mental functioning beyond chromosomal sex. Nevertheless, these positive results are all from the Baron-Cohen group.

Kung et al. (2016) is an external source of replication and neither replicates the CAST sex differences nor its relationship with FT. The combined sample of the two studies providing affirmative results were $N = 364$ compared to the $N = 92$ in which the results could not be replicated. Results from the first two datasets are favoured due to the lesser margin of error in the results (Tanur, 2011).

Nevertheless, whereas the articles with positive results received higher STROBE ratings than Kung et al. (2016) (77.3% and 81.8% relative to 63.6%), the latter retained a higher proportion of respondents (86% versus 45.6% and 51.9%). Though the findings that support EMB had more power, and a lower risk of bias, the null results may be more representative of the typical offspring from mothers who received amniocentesis.

FT on sex differences in ASD-related cognitive styles

Within the seven papers that investigated ASD-related cognition, two investigated cognitive styles (Finegan et al., 1992; Auyeung et al., 2012). Sex differences were not detected on any of the dimensions, except for the male superiority in the Embedded Figures Task (EFT) and in mental rotation (Auyeung et al., 2012). FT did not correlate with mental rotation performance and was a significant predictor of EFT performance above and beyond sex. The present findings support FT in mediating sex differences in EFT

In the four focused on hemispheric lateralization in language and emotional word processing (Beking et al., 2018; Grimshaw et al., 1995; Lust et al., 2010, 2011), sex differences were found in two domains: 1) girls showed larger vocabularies than males at 12th and 24th month (Lutchmaya et al., 2001); 2) boys had a right ear advantage in dichotic listening compared to the left ear, whereas this was not found in girls (Lust et al., 2010) and 3) mental rotation was more lateralized in the boys' right hemispheres than girls (Beking et al., 2018). FT accounted for the former two cognitive domains beyond sex, though the pattern of lateralization was inconsistent across studies (Lust et al., 2011). In Beking et al. (2018), FT acted as a moderator to determine the direction of association of pubertal testosterone and spatial cognition scores.

Systemizing and restricted interest were measured in two separate studies (Auyeung, 2006; Knickmeyer et al., 2005). Both constructs showed sex differences. FT contributes to high levels of these traits even after controlling for gender. Their STROBE scores were 77.3% and 63.6%.

Among the four studies that supported the role of FT affecting ASD-related cognition, three were from the Baron-Cohen group (75%).

FT on sex differences in social functioning

All social measures (Chapman et al., 2006; Knickmeyer et al., 2005, 2006; Lutchmaya et al., 2002) showed sex differences, except for the Reading the Mind in the Eyes task and social relationship quality. Of such findings, FT predicted mental, affective and neutral propositions, and eye-contact frequency beyond sex (Knickmeyer et al., 2006; Lutchmaya et al., 2002) FT inversely associated with relationship quality (Lutchmaya et al., 2002). STROBE ratings varied from 63.6% to 83.8%. All three studies were conducted by the Baron-Cohen group.

FT on sex differences in play and behaviour

Among the studies that examined play and other loci of behaviour (Grimshaw et al., 1995; Knickmeyer, Wheelright et al., 2005; Auyeung et al., 2009b; Körner et al., 2019; Spencer et al., 2017; van de Beek et al., 2009), sex differences were found across all indices. FT accounted for the variance of play in Auyeung et al. (2009b), with sex being the stronger predictor ($\beta = .77$) versus ft ($\beta = .14$). In Körner et al. (2019), FT predicted Delay of Gratification at $r = .23$ in boys, and did not significantly correlate with the measure in girls. While Auyeung et al. (2009b) was the largest of the four studies ($N = 212$), the other studies had the power to detect a within gender effect but did not find it. STROBE ratings were comparable between the two studies with positive (54.6–68.2%) versus three reporting null results (63.6–83.8%). No significant imbalance of independent investigators was detected here (see Bias Testing).

Age and overall outcome

Analysis of positive results by child's age, among those who reported this variable, did not indicate its effect as a moderator. Four studies of five (80%) that looked at infants (age 0 to 3 years) (Lutchmaya et al., 2001, 2002; Auyeung et al., 2010), three of five (60%) that looked at early childhood (3 to 6 years) (Knickmeyer et al., 2005, 2006), four out of eleven (36%) in middle-to-late childhood (6 to 10 years) and one out of one (100%) (Beking et al., 2018) in early adolescents (10 + years) showed results congruent with the EMB (Auyeung 2006; Auyeung et al., 2009a, 2009b; Lust et al., 2010). Pearson's Chi-Square showed no significant difference in positive results across age groups ($\chi^2(3, N = 22) = 3.67, P > .05$).

Bias testing

Chi-square independence test indicate that research group from which the data was obtained significantly related to the degree to which the results supported the EMB ($\chi^2(2, N = 22) = 4.46, P < .05$). Baron-Cohen's research group produced 75% of the positive results ($N = 9$), significantly more from other sources ($Z = 2.4, P < .05$).

The same analysis conducted on ASD symptoms, ASD-related cognition, sociality and playstyles yielded these results. Respectively, they are ($\chi^2(1, N = 4) = 4, P < .05$, 75%, $N = 3$), ($\chi^2(1, N = 9) = 1.10, P < .05$, 75%, $N = 3$) and, ($\chi^2(1, N = 3) = O.33, P > .05$) and, ($\chi^2(1, N = 6) = 0.38, P > .05$)

Homogeneity of sampling

Among the 12 articles with positive results, 10 (83%) were derived from projects that had produced EMB-congruent data in at least one other study. Three out of three positive results (100%) focusing on ASD symptoms, five out of five (100%) on ASD-related cognition, one out of two (50%) on sociality and one out of two (50%) on playstyles were from subject pools that yielded more than one supportive findings. Positive results were not significantly associated with sample homogeneity ($\chi^2(3, N = 12) = 4.8, P > .05$).

Discussion

Summary

EMB's predictions were consistent with the pattern of data but conclusions remain preliminary until further testing. Whereas FT accounted for ASD symptoms beyond sex in most of the samples examined, null findings are predominant vis-à-vis ratings of playstyles. The most consistently positive results are those of FT mediating the cognitive sex disparities in physical versus social objects of attention; that is, FT is enabling for narrow interests and systemizing, and suppresses the frequency of eye-contact, the use of intentional propositions and relationship quality. In both cases, the effects of FT outstrip the influence of sex. These data dovetail with the highly significant signal in the adult sex differences literature whereby men and women respectively displayed large divergences in preference toward thing and people-related vocations ($d = .93$) (Su, Rounds, & Armstrong, 2009). Together, FT emerges as a factor that mainly differentiates the sexes in cognition and behaviour that derive from motivation toward objects versus subjects. Psychometric data further support ASD symptoms as being partly accounted for by reduced social interests (Xiong, unpublished manuscript).

The conflicting results in accounting for ASD-related traits notwithstanding, the weight of evidence agrees with EMB. Though ASD is validated here as a manifestation of the male endophenotype, researchers should be cautious regarding the generality of this claim. FT apparently affects the mind in a sex-dependent manner, so that females exposed to higher FT are not necessarily more masculine, per se, as was the case where FT was associated with language comprehension and conceptual grouping, which did not show sex differences (Finegan et al., 1992). As demonstrated in Positron Emission Tomography (PET) studies of mathematical problem solving, equally competent performance of identical tasks recruits distinct brain circuitry in males than in females (Haier, White, & Alkire, 2003; Halpern et al., 2007). Whereas neuroimaging studies have demonstrated structural similarities of ASD females with neurotypical males (Ecker & Murphy, 2014), rather than a uniform masculinization across the sexes, feminization and masculinization of connectivity patterns are respectively found in male and female ASD patients (Alaerts, Swinnen, & Wenderoth, 2016). The data suggest that sex-specific neurophysiological milieus can interact with FT to express ASD phenotypes in a manner contingent upon whether the child is male or female.

This understanding of FT as a causal agent converges with the results from relevant animal and human studies. In-utero injection of testosterone led to behavioral masculinization in female macaques, suggesting a causal role of FT in primates for male-typical traits (Goy, Bercovitch & Brair, 1988). Behavioral continuity of human infants exists from the womb to after birth, as mothers retroactively report greater male activity in the 4th, 7th, 8th, and 9th month of gestation. Prenatal activity in turn predicts post-natal behavioral levels (Ellis & He, 2014), which suggest that the process of masculinization can be traced back to the second trimester. Though there are no randomized controlled trials involving the administration of testosterone that definitively demonstrate causality, the current review suggests that the temporal direction of FT is consistent with such a picture.

Null findings include aggressiveness and playstyles that are not consistently accounted for by FT. Where the link exists between FT and gendered play, the relationship is likely an artefact of sex. Sex-typical playstyles in children emerge as early as 9 to 12 months, preceding the cognitive maturity at 24 months which enables them to report on their own gender (Maccoby, 1998). The latter is necessary for modelling of same-sex peers. While FT levels at 12–24 weeks of gestation is not supported to be the mechanism that differentiates play behaviour between boys and girls, this does not necessarily rule out biology. Low second-to-fourth digit ratios are associated with masculinized play. If this can be accepted as an index of high testosterone levels early in foetal development, then it would highlight that the timing of exposure in utero moderates the effect of FT levels on later personality (Hönekopp & Thierfelder, 2009).

Subgroup moderation analysis by age did not reveal significant results. Whereas 80% of samples from 0 to 3 years old reached criteria for supporting the EMB, this is the case with 3–6, 6–10 and 10+ children at 60%, 37% and 100%, respectively. Albeit underpowered, the current review did not support Su et al. (2009) meta-analysis finding that sex differences in thing-person interests decline with increasing age.

Limitations

The systematic review reported here assesses the direction of effect across FT studies, in reference to the EMB theory. Its main limitation, as with many sex differences reviews, is the variety of outcome measures across studies. Whereas the consideration of EMB encompasses many aspects of sex dimorphic endophenotypes, the dispersal of data across multiple measures raises the possibility of the 1) inadequate samples devoted for each facet and 2) using variables that are statistically redundant. Attempts to reproduce the results in a clear, unified and coherent manner would be enhanced via multivariate analyses. The lexical model of individual differences is a candidate psychometric strategy that supplies a necessarily exhaustive and robust taxonomy of discrete psychological processes (Eysenck, 1991; Goldberg, 1990). It consists of five factor-analytically derived constructs – Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism – collapsed from all value-neutral personality adjectives in the English semantic

network. The statistical structures are stable across demographic and linguistic divides (Eysenck, 1992; Vecchione, Alessandri, Barbaranelli, & Caprara, 2011; Zheng et al., 2008) and accounts for large, if not all, of the variance in novel self-report scales, such as grit, emotional intelligence, self-esteem, and sensation seeking (Aluja, García, & García, 2003; Credé, Tynan, & Harms, 2017; Lucas, Diener, & Suh, 1996; Van Rooy & Viswesvaran, 2004). A multi-trait consideration with these lexical constructs in combination with amniocentesis measures will raise the probability of identifying robust and discrete dimensions of sex differences that is acted upon by FT.

Another issue in need of redress is for replication by bodies of researchers independent of Baron-Cohen. So far, they have generated 75% of the positive findings. Homogeneity of such results from the same longitudinal amniocentesis projects, though frequently supplying multiple papers, was not statistically substantiated, stemming from low power due to the small number of studies or otherwise. Works of this method cannot be dismissed based upon its difficulty. Its scholarly imperative hearkens back to seminal studies aimed reconciling nativist and constructivist perspectives on sex (e.g. Jacklin & Maccoby, 1978; Maccoby & Jacklin, 1978). Whereas the EMB has drawn interest among researchers across psychology, insufficient attention has been devoted to directly assaying FT. Amniocentesis is becoming a procedure that is deployed more selectively. Commonly conducted in expectant mothers over the age of 30, it is associated with increased risk of miscarriage and physiological complications for mother and child (Beta, Lesmes-HereDia, Bedetti, & Akolekar, 2018; Wilson et al., 1998). This expectedly narrows the pool of FT data and calls for resourceful researchers to compensate for the lack. Newer and less invasive prenatal genetic testing are becoming accessible, bypassing the possibility of assaying fetal hormones altogether (Ohno & Caughey, 2013). Nevertheless, amniocentesis is still the choice method in accurately assessing FT concentrations.

Conclusions

The central tenets of EMB has not been seriously challenged by empirical research. The shortage of replication studies from independent researchers render it uncertain whether the best measure of assessing amniotic hormones, short of extracting foetal blood, demonstrates FT as driving ASD. Recent data has in turn provided support for EMB. The neurophysiological portrait of those within the ASD classification also resembles that of males. Machine learning of neurotypical male and female brains by cortical thickness led to 80% of ASD females to be misclassified as males (Ecker et al., 2017). This dovetails with the interpretation of prospective data that FT masculinizes the brain in utero to accentuate ASD symptomatology and encourages further replication involving hormonal assays, which is the most immediate concern. Much remains unknown, as it is not clear how FT extracted through amniocentesis reflects the conditions that would affect the prenatal brain (Fine, 2005). Demonstrating the mechanism by which FT acts on the brain would be further step in uncovering the next most proximate factor to the development of male-typical cognition and ASD.

Appendix A. Results from term to subject heading search

Rows	2017 Searches	Results
Database 1: Embase, PsycINFO; Ovid Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid Medline; and PsycARTICLES Full Text up to December 14th, 2017		
#1	Adolescent or child or children or youth or infant* or neonate*	8,218,637
#2	F?etal testosterone or prenatal testosterone or gestational testosterone or uterine testosterone or amniotic testosterone or f?etal androgen or prenatal androgen or gestational androgen or uterine androgen or amniotic androgen or f?etal hormone or prenatal hormone or gestational hormone or uterine hormone or amniotic hormone or amniocentesis or extreme male brain or prenatal sex hormones (Title only)	5255
#3	Boy* or girl* or sex difference$ or gender difference$ or sex differentiation or gender differentiation male* or female*	17,676,859
#4	Social skills or theory of mind or mentalizing or empath* or care or visuospatial ability or perspective taking or social cognition or cognitive skill or cognitive ability social relationship communication or social perception or vocabulary size or eye gaze or eye contact or systemizing or autistic trait* or cerebral lateralization or language lateralization or behavior or behaviour or play	10,020,656
#5	#1 AND #2 AND #3 AND #4	215

Rows	2019 Searches	Results
Database 1: Embase, PsycINFO; Ovid Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) ALL; and PsycARTICLES Full Text up to December 4th, 2019		
#1	Adolescent or child or children or youth or infant* or neonate*	8,498,640
#2	Foetal testosterone or prenatal testosterone or gestational testosterone or uterine testosterone or amniotic testosterone or foetal androgen or prenatal androgen or gestational androgen or uterine androgen or amniotic androgen or f?etal hormone or prenatal hormone or gestational hormone or uterine hormone or amniotic hormone or amniocentesis or extreme male brain or prenatal sex hormones (Title only)	4893
#3	Boy* or girl* or sex difference$ or gender difference$ or sex differentiation or gender differentiation male* or female*	19,103,095
#4	Social skills or theory of mind or mentalizing or empath* or care or visuospatial ability or perspective taking or social cognition or cognitive skill or cognitive ability social relationship communication or social perception or vocabulary size or eye gaze or eye contact or systemizing or autistic trait* or cerebral lateralization or language lateralization or behavior or behaviour or play	12,801,904
#5	#1 AND #2 AND #3 AND #4	218

Rows	2020 Search	Results
Web of Knowledge		
#1	Adolescent or child or children or youth or infant* or neonate*	
#2	Foetal testosterone or prenatal testosterone or gestational testosterone or uterine testosterone or amniotic testosterone or foetal androgen or prenatal androgen or gestational androgen or uterine androgen or amniotic androgen or f?etal hormone or prenatal hormone or gestational hormone or uterine hormone or amniotic hormone or amniocentesis or extreme male brain or prenatal sex hormones (Title only)	
#3	Boy* or girl* or sex difference$ or gender difference$ or sex differentiation or gender differentiation male* or female*	
#4	Social skills or theory of mind or mentalizing or empath* or care or visuospatial ability or perspective taking or social cognition or cognitive skill or cognitive ability social relationship communication or social perception or vocabulary size or eye gaze or eye contact or systemizing or autistic trait* or cerebral lateralization or language lateralization or behavior or behaviour or play	
#5	#1 AND #2 AND #3 AND #4	576

References

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

Alaerts, K., Swinnen, S. P., & Wenderoth, N. (2016). Sex differences in autism: A resting-state fMRI investigation of functional brain connectivity in males and females. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsw027.

Aluja, A., García, Ó., & García, L. F. (2003). Relationships among extraversion, openness to experience, and sensation seeking. *Personality and Individual Differences*. https://doi.org/10.1016/S0191-8869(02)00244-1.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K., & Hackett, G. (2009a). Fetal testosterone and autistic traits. *British Journal of Psychology, 100*(1), 1–22.

Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K., Hackett, G., et al. (2009b). Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. *Psychological science, 20*(2), 144–148.

Auyeung, B., Baron-Cohen, S., Chapman, E., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Foetal testosterone and the child systemizing quotient. *European Journal of Endocrinology, 155*(suppl 1), S123–S130.

Auyeung, B., Knickmeyer, R., Ashwin, E., Taylor, K., Hackett, G., & Baron-Cohen, S. (2012). Effects of fetal testosterone on visuospatial ability. *Archives of sexual behavior, 41*(3), 571–581.

Auyeung, B., Taylor, K., Hackett, G., & Baron-Cohen, S. (2010). Foetal testosterone and autistic traits in 18 to 24-month-old children. *Molecular autism, 1*(1), 11.

Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes" *Intelligence*. https://doi.org/10.1016/j.intell.2014.03.001.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(02)01904-6.

Baron-Cohen, S. (2005). Testing the extreme male brain (EMB) theory of autism: Let the data speak for themselves. *Cognitive Neuropsychiatry, 10*(1), 77–81.

Baron-Cohen, S., Cassidy, S., Auyeung, B., Allison, C., Achoukhi, M., Robertson, S., et al. (2014). Attenuation of typical sex differences in 800 adults with autism vs. 3,900 controls. *PloS one, 9*(7), Article e102251.

Beking, T., Geuze, R. H., Van Faassen, M., Kema, I. P., Kreukels, B. P. C., & Groothuis, T. G. G. (2018). Prenatal and pubertal testosterone affect brain lateralization. *Psychoneuroendocrinology, 88*, 78–91.

Beta, J., Lesmes-HereDia, C., Bedetti, C., & Akolekar, R. (2018). Risk of miscarriage following amniocentesis and chorionic villus sampling: A systematic review of the literature. *Minerva Ginecologica*. https://doi.org/10.23736/S0026-4784.17.04178-8.

Caldwell-Harris, C. L., & Jordan, C. J. (2014). Systemizing and special interests: Characterizing the continuum from neurotypical to autism spectrum disorder. *Learning and individual differences, 29*, 98–105.

Chakrabarti, S., & Fombonne, E. (2001). Pervasive developmental disorders in preschool children. *Jama, 285*(24), 3093–3099.

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the empathy quotient (EQ) and the "reading the mind in the eyes" test. *Social Neuroscience, 1*(2), 135–148.

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., de Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric Disease and Treatment*. https://doi.org/10.2147/NDT.S104620.

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000102.

Ecker, C., Andrews, D. S., Gudbrandsen, C. M., Marquand, A. F., Ginestet, C. E., Daly, E. M., et al. (2017). Association between the probability of autism spectrum disorder and normative sex-related phenotypic diversity in brain structure. *Jama psychiatry, 74*(4), 329–338.

Ellis, L., & He, P. (2014). Sex differences in fetal activity and childhood hyperactivity. *Research Journal of Developmental Biology, 1*(1), 1.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences*. https://doi.org/10.1016/0191-8869(92)90237-J.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3?-Criteria for a taxonomic paradigm. *Personality and Individual Differences*. https://doi.org/10.1016/0191-8869(91)90144-Z.

Fine, C. (2005). *Delusions of gender: The real science behind sex differences.* Icon Books Ltd.

Finegan, J. A. K., Niccols, G. A., & Sitarenios, G. (1992). Relations between prenatal testosterone levels and cognitive abilities at 4 years. *Developmental Psychology, 28*(6), 1075.

Finegan, J. A. K., Quarrington, B. J., Hughes, H. E., Rudd, N. L., Stevens, L. J., Weksberg, R., et al. (1984). Midtrimester amniocentesis: Obstetric outcome and neonatal neurobehavioral status. *American Journal of Obstetrics and Gynecology*. https://doi.org/10.1016/0002-9378(84)90396-X.

Geschwind, N., & Galaburda, A. M. (1985). Cerebral lateralization: Biological mechanisms, associations, and pathology: I. A hypothesis and a program for research. *Archives of Neurology, 42*(5), 428–459.

Gillberg, C., Cederlund, M., Lamberg, K., & Zeijlon, L. (2006). Brief report: "the autism epidemic". The registered prevalence of autism in a Swedish urban area. *Journal of Autism and Developmental Disorders, 36*(3), 429.

Goldberg, L. R. (1990). An alternative "Description of Personality": The big-five factor structure. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.59.6.1216.

Goy, R. W., Bercovitch, F. B., & McBrair, M. C. (1988). Behavioral masculinization is independent of genital masculinization in prenatally androgenized female rhesus macaques. *Hormones and behavior, 22*(4), 552–571.

Grimshaw, G. M., Bryden, M. P., & Finegan, J. A. K. (1995a). Relations between prenatal testosterone and cerebral lateralization in children. *Neuropsychology, 9*(1), 68.

Grimshaw, G. M., Sitarenios, G., & Finegan, J. A. K. (1995b). Mental rotation At 7 years-relations with prenatal testosterone levels and spatial play experiences. *Brain and Cognition, 29*(1), 85–100.

Haier, R. J., White, N. S., & Alkire, M. T. (2003). Individual differences in general intelligence correlate with brain function during nonreasoning tasks. *Intelligence*. https://doi.org/10.1016/S0160-2896(03)00025-4.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, Supplement.*. https://doi.org/10.1111/j.1529-1006.2007.00032.x.

Hines, M. (2006). Prenatal testosterone and gender-related behaviour. *European Journal of Endocrinology*. https://doi.org/10.1530/eje.1.02236.

Hönekopp, J., & Thierfelder, C. (2009). Relationships between digit ratio (2D: 4D) and sex-typed play behavior in pre-school children. *Personality and Individual Differences, 47*(7), 706–710.
Jacklin, C. N., & Maccoby, E. E. (1978). Social behavior at thirty-three months in same-sex and mixed-sex dyads. *Child Development*. https://doi.org/10.2307/1128222.
Jensen, P. S., Martin, D., & Cantwell, D. P. (1997). Comorbidity in ADHD: Implications for research, practice, and DSM-V. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(8), 1065–1079.
Knickmeyer, R., Baron-Cohen, S., Raggatt, P., & Taylor, K. (2005). Foetal testosterone, social relationships, and restricted interests in children. *Journal of Child Psychology and Psychiatry, 46*(2), 198–210.
Knickmeyer, R., Baron-Cohen, S., Raggatt, P., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy. *Hormones and Behavior, 49*(3), 282–292.
Knickmeyer, R. C., Wheelwright, S., Taylor, K., Raggatt, P., Hackett, G., & Baron-Cohen, S. (2005). Gender-typed play and amniotic testosterone. *Developmental Psychology, 41*(3), 517.
Körner, L. M., Pause, B. M., Meinlschmidt, G., Tegethoff, M., Fröhlich, S., Kozlowski, P., et al. (2019). Prenatal testosterone exposure is associated with delay of gratification and attention problems/overactive behavior in 3-year-old boys. *Psychoneuroendocrinology, 104*, 49–54. https://doi.org/10.1016/j.psyneuen.2019.02.014.
Kung, K. T., Spencer, D., Pasterski, V., Neufeld, S., Glover, V., O'connor, T. G., ... & Hines, M. (2016). No relationship between prenatal androgen exposure and autistic traits: convergent evidence from studies of children with congenital adrenal hyperplasia and of amniotic testosterone concentrations in typically developing children. Journal of Child Psychology and Psychiatry, 57(12), 1455–1462.
Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant Validity of Well-Being Measures. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.71.3.616.
Lust, J. M., Geuze, R. H., Van de Beek, C., Cohen-Kettenis, P. T., Bouma, A., & Groothuis, T. G. (2011). Differential effects of prenatal testosterone on lateralization of handedness and language. *Neuropsychology, 25*(5), 581.
Lust, J. M., Geuze, R. H., Van de Beek, C., Cohen-Kettenis, P. T., Groothuis, A. G. G., & Bouma, A. (2010). Sex specific effect of prenatal testosterone on language lateralization in children. *Neuropsychologia, 48*(2), 536–540.
Lutchmaya, S., Baron-Cohen, S., & Raggatt, P. (2001a). Foetal testosterone and vocabulary size in 18-and 24-month-old infants. *Infant Behavior and Development, 24*(4), 418–424.
Lutchmaya, S., Baron-Cohen, S., & Raggatt, P. (2002bc). Foetal testosterone and eye contact in 12-month-old human infants. *Infant Behavior and Development, 25*(3), 327–335.
Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., & Manning, J. T. (2004). 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Human Development, 77*(1), 23–28.
Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.
Maccoby, E. E., & Jacklin, C. N. (1978). The psychology of sex differences (Vol. 2). Stanford University Press.
Mathis, M. A. D., Alvarenga, P. D., Funaro, G., Torresan, R. C., Moraes, I., Torres, A. R., et al. (2011). Gender differences in obsessive-compulsive disorder: A literature review. *Revista Brasileira de Psiquiatria, 33*(4), 390–399.
McManus, I. C., & Bryden, M. P. (1991). Geschwind's theory of cerebral lateralization: Developing a formal, causal model. *Psychological bulletin, 110*(2), 237.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine, 6(7), e1000097.
Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-005-0040-7.
Ohno, M., & Caughey, A. (2013). The role of noninvasive prenatal testing as a diagnostic versus a screening tool - a cost-effectiveness analysis. *Prenatal Diagnosis*. https://doi.org/10.1002/pd.4156.
Rodeck, C. H., Gill, D., Rosenberg, D. A., & Collins, W. P. (1985). Testosterone levels in midtrimester maternal and fetal plasma and amniotic fluid. *Prenatal Diagnosis, 5*(3), 175–181.
Slijper, F. M. (1984). Androgens and gender role behaviour in girls with congenital adrenal hyperplasia (CAH). In Progress in brain research (Vol. 61, pp. 417-422). Elsevier.
Spencer, D., Pasterski, V., Neufeld, S., Glover, V., O'Connor, T. G., Hindmarsh, P. C., et al. (2017). Prenatal androgen exposure and children's aggressive behavior and activity level. *Hormones and Behavior*. https://doi.org/10.1016/j.yhbeh.2017.09.012.
Stevens, T., Peng, L., & Barnard-Brak, L. (2016). The comorbidity of ADHD in children diagnosed with autism spectrum disorder. *Research in Autism Spectrum Disorders*. https://doi.org/10.1016/j.rasd.2016.07.003.
Stodgell, C. J., Ingram, J. L., & Hyman, S. L. (2000). *The role of candidate genes in unraveling the genetics of autism.*. https://doi.org/10.1016/s0074-7750(00)80006-3.
Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and Things, Women and People: A Meta-Analysis of Sex Differences in Interests. Psychological Bulletin. Doi: 10.1037/a0017364.
Tanur, J. M. (2011). Margin of Error. In International Encyclopedia of Statistical Science (pp. 765-765). Springer Berlin Heidelberg.
Tavassoli, T., Miller, L. J., Schoen, S. A., Jo Brout, J., Sullivan, J., & Baron-Cohen, S. (2017). Sensory reactivity, empathizing and systemizing in autism spectrum conditions and sensory processing disorder. *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/J.DCN.2017.05.005.
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. *Archives of Sexual Behavior, 38*(1), 6–15.
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*. https://doi.org/10.1016/j.yhbeh.2004.06.010.
Vandenbroucke, J. P., Von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., et al. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *PLoS medicine, 4*(10), Article e297.
Van Rooy, D. L., & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior*. https://doi.org/10.1016/S0001-8791(03)00076-9.
Vecchione, M., Alessandri, G., Barbaranelli, C., & Caprara, G. (2011). Higher-order factors of the big five and basic values: Empirical and theoretical relations. *British Journal of Psychology*. https://doi.org/10.1111/j.2044-8295.2010.02006.x.
Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D: 4D) and Baron-Cohen's "Reading the Mind in the Eyes" test, empathy, systemising, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences, 41*(8), 1481–1491.
Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. Frontiers in psychology, 2.
Whitehouse, A. J. (2016). Commentary: Are we expecting too much from the extreme male brain theory of autism? A reflection on Kung et al.(2016). Journal of Child Psychology and Psychiatry, 57(12), 1463-1464.
Wing, L. (1985). *Autistic children: A guide for parents and professionals.* New York: Brunner/Mazel.
Wilson, R. D., Johnson, J. M., Dansereau, J., Singer, J., Drinnan, S. L., Winsor, E. J. T., et al. (1998). Randomised trial to assess safety and fetal outcome of early and midtrimester amniocentesis. *Lancet*. https://doi.org/10.1016/S0140-6736(97)12346-7.
Xiong, H., Swift, V. and Peterson, J. B. (May 2016) Modelling Systemizing and Empathizing with the Big Five Aspects: Implications for the Extreme Female Brain Theory of Psychosis. Poster: American Psychological Society, Chicago, Illinois.
Yang, Y., & Barth, J. M. (2015). Gender differences in STEM undergraduates' vocational interests: People–thing orientation and goal affordances. *Journal of Vocational Behavior, 91*, 65–75.
Zheng, L., Goldberg, L. R., Zheng, Y., Zhao, Y., Tang, Y., & Liu, L. (2008). Reliability and concurrent validation of the IPIP Big-Five factor markers in China: Consistencies in factor structure between Internet-obtained heterosexual and homosexual samples. *Personality and Individual Differences*. https://doi.org/10.1016/j.paid.2008.07.009.