



Dump the “dimorphism”: Comprehensive synthesis of human brain studies reveals few male-female differences beyond size



Lise Eliot ^{a,b,*}, Adnan Ahmed ^b, Hiba Khan ^b, Julie Patel ^b

^a Department of Foundational Sciences and Humanities, Neuroscience Discipline, Rosalind Franklin University of Medicine & Science, 3333 Green Bay Rd., North Chicago, IL, 60064, USA

^b Chicago Medical School, Rosalind Franklin University of Medicine & Science, 3333 Green Bay Rd., North Chicago, IL, 60064, USA

ARTICLE INFO

Keywords:

Sex
Gender
MRI
Meta-analysis
Corpus callosum
Hippocampus
Amygdala
Massa intermedia
Anterior commissure
Lateralization
Connectome
Default mode network
Verbal
Spatial
Mental rotation
Emotion
Empathy
Precision medicine
Multivariate
Cortical thickness

ABSTRACT

With the explosion of neuroimaging, differences between male and female brains have been exhaustively analyzed. Here we synthesize three decades of human MRI and postmortem data, emphasizing meta-analyses and other large studies, which collectively reveal few reliable sex/gender differences and a history of unreplicated claims. Males' brains are larger than females' from birth, stabilizing around 11 % in adults. This size difference accounts for other reproducible findings: higher white/gray matter ratio, intra- versus interhemispheric connectivity, and regional cortical and subcortical volumes in males. But when structural and lateralization differences are present independent of size, sex/gender explains only about 1% of total variance. Connectome differences and multivariate sex/gender prediction are largely based on brain size, and perform poorly across diverse populations. Task-based fMRI has especially failed to find reproducible activation differences between men and women in verbal, spatial or emotion processing due to high rates of false discovery. Overall, male/female brain differences appear trivial and population-specific. The human brain is not “sexually dimorphic.”

1. Introduction

How different are male and female brains? Sex and gender are widely regarded as critical variables in human neuroscience. Although it is customary to refer to “sex” as the inflexible biological component of male-female difference (e.g. chromosomes, reproductive organs, and gonadal hormone titers) and “gender” as the psychosocial manifestation of human male- and femaleness (e.g., identity, behavior, appearance, social role; [National Institutes of Health, 2015](#)), both phenotypes are deeply entangled when it comes to the brain and the neural processes giving rise to masculinity and femininity. Given abundant evidence that experience alters neuronal structure and function, as well as growing

knowledge about epigenetic influences on CNS development, it is impossible to discern the degree to which group-level differences between human males and females are attributable to inborn sex factors versus social-environmental gender learning, acting through lifelong neuroplasticity. Reflecting this bidirectional entanglement, we here adopt the term “sex/gender” (s/g) to label the independent variable that is captured in brain studies using a mixed population of male and female participants ([Jordan-Young and Rumiati, 2012](#); [Kaiser et al., 2009](#); [Springer et al., 2012](#)), where group assignment is nearly always based on self-identity. Another rationale for using the conjunctive label is that they are indexed as a single term in PubMed.

Whatever we call it, there is no question that s/g contributes to

* Corresponding author at: Department of Foundational Sciences and Humanities, Neuroscience Discipline, Rosalind Franklin University of Medicine & Science, 3333 Green Bay Rd., North Chicago, IL, 60064, USA.

E-mail address: lise.eliot@rosalindfranklin.edu (L. Eliot).

individual variance in neurobehavioral research. Many psychiatric and neurological disorders exhibit unequal prevalence between males and females (Fig. 1), leading many to assume they are caused by preexisting s/g differences in the brain. This is the predominant view put forth in prior narrative reviews of this literature, which have assembled findings of brain differences between males and females at levels ranging from the molecular (Jazin and Cahill, 2010) to structural and functional MRI (Cahill, 2006; Cosgrove et al., 2007; Giedd et al., 2012; Grabowska, 2017; Hines, 2011; Kaczurkin et al., 2019). Such studies complement decades of research in behavioral science, where s/g-related differences have been extensively documented and quantified (Halpern, 2012; Hyde, 2016; Lippa, 2002; Zell et al., 2015).

However, unlike the more mature field of psychology, where s/g findings have been subjected to comprehensive analyses, s/g difference in the brain has only rarely been synthesized in a systematic manner. In psychology, large-scale meta-analyses and meta-syntheses of the literature have found that human males and females are far more similar than different in most measures of cognition, personality and attitudes (Carothers and Reis, 2013; Hyde, 2005; Zell et al., 2015). Except for a few behaviors such as physical aggression, mental rotation ability, and peer attachment, some 85 % of s/g differences exhibit effect sizes smaller than $d = 0.35$, and thus considered “small” by Cohen’s criterion (Zell et al., 2015). In neuroscience, only a few meta-analyses have thus far examined s/g difference in the human brain, but the findings are similar: much more similarity than difference between males and females in measures of brain structure (Bishop and Wahlsten, 1997; Marwaha et al., 2017; Ruigrok et al., 2014; Tan et al., 2016) and function (Garcia-Garcia et al., 2016; Sergerie et al., 2008; Sommer et al., 2008; Yuan et al., 2019).

In the present synthesis, we have taken special effort to capture a large number of studies that use similar methodology to compare males and females in specific measures of brain structure and function. Whenever available, we use findings from meta-analyses or systematic reviews. For other structural and functional imaging findings that have yet to be systematically reviewed, we have tabulated summaries of articles identified through repeated PubMed and Google Scholar searches, including the largest and most highly-cited studies. Given the wide range of brain measures we have included in this single review, none of our searches have been fully systematic. Nonetheless, we have aimed throughout to integrate best practices for systematic review (Siddaway et al., 2019) within a historical framework that avoids vote-counting to provide a narrative meta-synthesis of the current state of the field.

A key strategy in searching the literature for this topic is to look beyond the terms “sex difference” and “gender difference,” which explicitly select for positive findings and are likely to miss studies that

disaggregated male and female data but did not find a difference between them. Like most areas of research, s/g-related brain findings are vulnerable to reporting bias, wherein analyses that find a positive difference between the brains of men and women are likelier to be published than those that failed to find a significant difference. This bias against publishing negative data, known as the “file drawer effect” (Rosenthal, 1979), is coming under increasing scrutiny through systematic analysis and statistical proof of “excess significance” in the literature on brain abnormalities (Ioannidis, 2011) and cognitive science (Ioannidis et al., 2014). Such analyses assume that multiple replications of the same study should produce a normal distribution of effect sizes around the pooled, meta-analytic mean. When the collection of published effect sizes instead forms a skewed distribution, this provides evidence of excess publication of larger, statistically significant findings and conspicuous absence of smaller and non-significant findings. For comparisons of males and females, this bias leans toward the under-publication of similarity (Song et al., 2010) and relative over-reporting of data that finds statistically significant s/g difference (Halpern, 2012, p. 76), perhaps reflecting the high public salience of s/g difference findings (Eliot, 2011; Maney, 2015; Rippon, 2016).

Direct evidence for publication bias comes from a recent meta-analysis of sex differences identified in 179 fMRI studies of brain activation (David et al., 2018). This analysis found no correlation between study size and the number of brain foci each study identified as showing statistically-significant s/g differences in activity. This is contra the prediction that larger studies should detect a greater number of significant foci due to their greater statistical power. In addition, only 12 % of identified studies stated null s/g findings in their abstract, an “implausibly high” success rate given the overall distribution of findings. David et al. (2018) concluded that the fMRI literature on brain sex difference is afflicted by reporting bias and excess significance, in part because post-hoc “subgroup analyses based on sex are always tempting to do and easy to perform.”

Indeed, reporting bias has been implicated in other research generated through post-hoc comparison of male and female groups from larger preexisting datasets. Thus, an empirical assessment of 432 sex difference claims from 77 different gene-disease association studies found that the vast majority of such claims were “spurious” or inadequately documented (Patsopoulos et al., 2007), concluding that selective reporting has led to an excess of positive s/g effects in the published literature. Similar results emerged from a synthesis of randomized clinical trials by Wallach et al. (2016), which found that s/g treatment interactions were no more common than would be predicted by chance and were largely uncorroborated by meta-analysis. In other words, most appear to be false positives.

Mindful of such reporting biases, we sought to integrate findings from a wide survey of the extant literature on s/g effects in human structural and functional brain imaging, using the search terms “sex” and “gender” in lieu of “sex difference” and “gender difference.” Our analysis confirms large group-level differences between men and women in one particular measure: overall brain volume. However, once individual differences in total brain size are accounted for, there is scant evidence of reliable s/g-specific enhancement of any particular brain structure or activity pattern. Male and female brains are overwhelmingly similar, in contrast to the male-female differences in reproductive organs that can properly be called “sexually dimorphic” (Joel, 2012).

2. Brain size, sexual dimorphism, and the methodology of male-female comparison

The most obvious difference between male and female brains is in overall mass or volume. We thus begin with a discussion of global brain size and how this affects the various methods used to compare between males and females when deciding whether particular brain structures or connectivity patterns are “sexually dimorphic.” Whenever possible, we

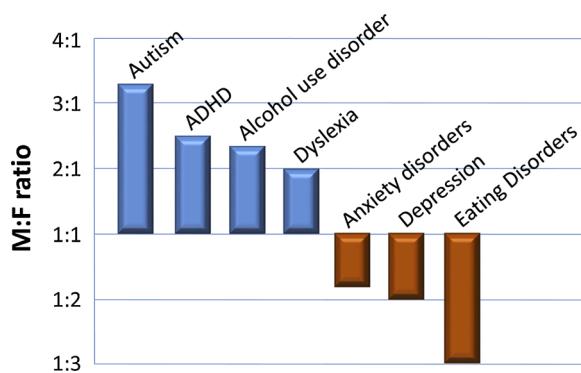


Fig. 1. Male:Female (M:F) prevalence ratios for common behavioral disorders.

SOURCES: Autism (Loomees et al., 2017); ADHD (Polanczyk et al., 2007); alcohol use disorder (Grant et al., 2004); dyslexia (Rutter et al., 2004); anxiety disorders (McLean et al., 2011); depression (Salk et al., 2017); eating disorders (Hudson et al., 2007).

quantify s/g differences by the more intuitive “percent difference” as well as by effect size, or d -value (the difference between male and female means divided by the population standard deviation), using the convention of Cohen (1992) for labeling their magnitude (0.2 = small; 0.5 = moderate; 0.8 = large).

2.1. Sex/gender difference in global brain size

Research over many decades has found that global brain size is unambiguously larger in males. This is usually reported as either total intracranial volume (ICV; often also abbreviated to TIV), or as total brain volume (TBV) which excludes the meninges and ventricles. In a 2014 meta-analysis involving 77 studies and nearly 15,000 participants, Ruigrok et al. found that ICV is 12 % larger in males, a highly significant and very large difference with an effect size of $d = 3.0$. They further found cerebrospinal fluid (CSF) volume to be 11.5 % larger in males ($d = 1.2$) and TBV to be 10.8 % larger in males ($d = 2.1$), based on a subset of studies that included 4,484 and 2,532 participants, respectively. Since TBV excludes the CSF compartments (ventricles and subarachnoid space), it stands to reason that the s/g difference is smaller for TBV than for ICV. A nearly identical difference in TBV (10.6 %) was reported in a large UK Biobank study (5,216 participants) published more recently (Ritchie et al., 2018).

S/g difference in global brain volumes are present at all ages, although studies of infants and children are rarer and generally much smaller than studies of adults. Ruigrok et al. (2014) found larger ICV in males from birth onward, with the difference appearing to peak in adolescence. In another review, Paus (2010) reported global brain volume as about 8% larger in male neonates, increasing to a peak of roughly 13 % during adolescence/early adulthood. More recently, Knickmeyer et al. (2017) imaged a large sample of neonates and found that ICV, total gray matter and total white matter are each about 5% larger in males, while Dean et al. (2018) reported a difference of 8% at one-month of age. Thus, the s/g difference in TBV is clearly present from birth and increases in magnitude during postnatal development.

Larger bodies require larger brains and the s/g difference in brain volume mostly parallels the divergence of male/female body size during development. Thus, newborn boys are about 4% heavier (Talge et al., 2014) than newborn girls, with the difference increasing to an average of 18 % greater weight and 9% greater height in adult males (McDowell et al., 2008). Longitudinal imaging of 387 adolescents found that the s/g difference in TBV grew steadily from a 6% difference at age 7 to a 15 % difference at age 20, in parallel with the divergence in height and weight over adolescence (Lenroot et al., 2007; Paus et al., 2017). Brain weight and height are correlated similarly in both sexes (Forstmeier, 2011). Nonetheless, it is notable that the 11 % brain volume difference between men and women is smaller than the s/g difference in other internal organs, including the heart (17 %), lung (23 %), liver (14 %), pancreas (18 %), kidney (19 %), and thyroid (25 %; de la Grandmaison et al., 2001). Limited data in children similarly indicate that like the brain, other internal organs are larger in boys even before puberty (kidney: Schmidt et al., 2001; heart: Janz et al., 2000).

With larger brains come larger raw volumes of each CNS compartment in males: gray matter (GM), white matter (WM), and ventricular volumes (Brain Development Cooperative Group, 2012; Chen et al., 2007; Krugel, 2006; Lemaître et al., 2005; Lenroot et al., 2007; Paus, 2010; Pintzka et al., 2015; Ritchie et al., 2018; Tomasi and Volkow, 2012a). However, there are subtle differences in the way each compartment varies by s/g, with most studies reporting a higher ratio of GM-to-WM in females (Chen et al., 2007; Cosgrove et al., 2007; Luders and Toga, 2010). The magnitude of this difference ranged from 4 to 7% across six studies reviewed by Leonard et al. (2008), in agreement with a 5.5 % greater female GM/WM in the large study by Ritchie et al. (2018) and 6% in Pintzka et al. (2015). Importantly, Leonard et al. (2008) and Luders et al. (2002) both demonstrated that GM/WM ratio is itself a function of brain size, with Jäncke et al. (2015) reporting the GM/WM

s/g difference to be “massively reduced” (to roughly $d = 0.05$) when covaried against TBV. Thus, larger brains have a higher proportion of WM than smaller brains, regardless of sex or even species (de Jong et al., 2017). This non-uniform scaling reflects the need for more (or larger diameter, or more heavily myelinated) axons to faithfully transmit across a greater physical distance (Bush and Allman, 2003; Zhang and Sejnowski, 2000). Accordingly, men and women matched for brain size were found to exhibit no difference in GM/WM ratio (Luders et al., 2009). Moreover, the GM/WM ratio is found to diverge the most after puberty, when sex differences in height and total brain size reach their peak (Lenroot et al., 2007; Paus, 2010). This non-linear scaling of WM to brain size is important when comparing the structural “connectome” between males and females (see Section 5).

2.2. Are specific brain structures larger in men or women? Origin of the term “sexual dimorphism” in the CNS

Statistically-speaking the ~11 % male/female difference in TBV is a large effect ($d > 0.80$). But this is not the kind of “sexual dimorphism” typically of interest to neuroscientists. Rather, the search for human brain sexual dimorphisms has been based on the supposition that *specific* structures or circuits differ disproportionately between men and women in ways that will explain well-known behavioral s/g differences, such as empathy, spatial navigation, and gender identity itself. This paradigm has its roots in animal neurobiology, where certain brain areas are indeed dramatically larger in one sex, with clear links to behaviors such as courtship and mating. Thus, Nottebohm and Arnold (1976) first found that vocal control areas in the brain of canaries and zebra finches are as much as 6-fold larger in male birds, species in which only the males normally sing. Another striking example is the spinal nucleus of the bulbocavernosus (SNB) in rodents, which innervates two muscles at the base of the penis and is clearly present in males’ lumbar spinal cord but barely visible in adult females (Breedlove and Arnold, 1980). Yet another structure is literally called the “sexually-dimorphic nucleus” (SDN) and is located in the anterior hypothalamus where it can measure up to 5-fold larger in male rats, compared to females (Gorski et al., 1978).

More recent research has challenged the notion that such dimorphic structures necessarily mediate sexually-differentiated behavior (De Vries, 2004; Ball, 2016). Nonetheless, this paradigm drives much of the search for brain s/g differences underlying human behavioral differences. The human brain lacks song control nuclei, but the homologue of the SNB exists as a portion of Onuf’s nucleus in the lumbar spinal cord, a structure that is not orders-of-magnitude larger (as in the rat), but about one-third larger in men than women (Purves et al., 2018, pp. 677–78). In the case of the SDN, the search for its human homologue took nearly 20 years to reach consensus, but was finally settled upon as the third interstitial nucleus of the anterior hypothalamus (INAH-3), a tiny (0.1 mm^3) subnucleus situated lateral to the much larger medial preoptic nucleus. Again, the magnitude of this difference is a fraction of the 5-fold rodent difference. Four different labs reported that the structure is larger in men, but the difference averages only 1.6-fold (Allen et al., 1989; Byne et al., 2000; Garcia-Falgueras and Swaab, 2008; LeVay, 1991). Nor is there a clear relationship of INAH-3 volume to sexual behavior: LeVay (1991) reported that the structure is smaller in homosexual, compared to heterosexual men, whereas Byne et al. (2001) found no significant difference between such groups. With regard to gender identity, Garcia-Falgueras and Swaab (2008) reported reduced INAH-3 volume in a small sample of transgender women, but this has yet to be independently confirmed.

Small though it is, the reason INAH-3 has been so extensively studied is because this 60 % volume difference is by far the largest “sexual dimorphism” in the human brain. Nonetheless, the term is liberally applied to far more subtle male/female differences in human brain structure, neural activity, and even behavior. Moreover, since the classic sexual dimorphisms in rodent and songbird brains were found to be

influenced by early testosterone exposure (Arnold and Gorski, 1984), it is often assumed that any structural or functional differences between men and women's brains are the product of gonadal hormones acting prenatally and/or post-puberty. Thus, one of the most influential MRI studies in this field (1,078 citations), titled "Normal sexual dimorphism of the adult human brain," took pains to link specific structural volumes to androgen and estrogen receptor distributions from animal brains, even though many of the human volumes were not significantly different between the men and women in the study (Goldstein et al., 2001).

Thus, the common framing of human brains as "sexually dimorphic" is based on the model of X and Y chromosomes acting early in development and largely by way of gonadal hormones to enhance or suppress the growth of specific structures, essentially bifurcating male and female brains into distinct forms (Arnold, 2004). This binary classification has been widely extended to describe male-female neurophysiological or behavioral differences using the same adjective, "dimorphic" (e.g., Davis and Pfaff, 2014), even when the distribution of measures may be largely overlapping (Joel, 2011) and despite the caution urged by some in the field (McCarthy et al., 2012). But as the remainder of this paper will demonstrate, such binary classification does not accord with actual measures of human brain s/g difference, which are generally small, unreliable, and insignificant once individual body size is accounted for.

2.3. Scaling issues in the comparison of male and female brains

From the beginning, the search for sexual dimorphisms in the human brain has been faced with a scaling problem. Recognizing that brain size is related to body size and that human bodies are indeed sexually dimorphic, neuroscientists have struggled to find ways of comparing brain structures between men and women that don't merely reflect bodily size differences. Thus, every major study confirms that it is not only global measures like ICV, TBV, total GM, total WM and CSF volume that are larger in males, but every zone of the cerebral cortex and every subcortical structure, when reported as "raw," "native," or "uncorrected" volume (e.g., Brain Development Cooperative Group, 2012; Chen et al., 2007; Fjell et al., 2009; Goldstein et al., 2001; Lenroot et al., 2007; Luders et al., 2002; Nopoulos et al., 2000; Paus, 2010; Pintzka et al., 2015; Ritchie et al., 2018; Tomasi and Volkow, 2012a; Voevodskaya et al., 2014). To cope with individual differences in body size, brain imagers have developed various methods for normalizing regional volumes—a correction that is equally important whether the study is focused on sex, age, genotype, ethnicity, diagnosis, or any other grouping of interest.

In a few early studies (e.g., Raz et al., 2004), individuals' height was used as a covariate. However, height correlates poorly with both global and regional brain measures in a mixed-sex population (Jäncke et al., 2015; O'Brien et al., 2011; Peters et al., 1998; Witelson et al., 2006). Better correlation is found for measures within the head itself, such that most studies now normalize using either ICV or TBV.

Of the two measures, it was initially easier to manually outline ICV, so this became the more widely-used standard. As early as 1989, Clifford Jack and colleagues developed a regression or covariance method to calculate normalized regional brain volumes, adjusting the raw volume of a given structure by an amount proportional to the difference between a subject's ICV and the mean ICV for all subjects (Jack et al., 1989). The other common method is to take a simple ratio between individuals' regional brain volumes and their ICV (e.g., Goldstein et al., 2001). Both methods were found to reduce the variance in structural measures across the sample, demonstrating the importance of brain size normalization for understanding demographic- and disease-related volumetry (Free et al., 1995). However, as this and more recent studies (Barnes et al., 2010; O'Brien et al., 2011; Sanchis-Segura et al., 2019) have shown, covariance is superior to the ratio (also called "proportions") method in being less sensitive to distortion and measurement error, thereby retaining greater statistical power for group comparisons (Sanfilipo

et al., 2004). Of note, studies using a power-proportionate method, as opposed to linear proportion, also perform comparably to ICV covariance in reducing structural scaling errors (Liu et al., 2014; de Jong et al., 2017).

In recent work, Sanchis-Segura et al. (2019) demonstrated that ICV covariance (or regression) and the power-proportionate method were both superior to a linear proportion (ratio) method for group comparisons, judging by elimination of the correlation between global and normalized regional measures. Applying this to s/g differences, four studies (Jäncke et al., 2015; Pintzka et al., 2015; Sanchis-Segura et al., 2019; Voevodskaya et al., 2014) have now demonstrated that the linear proportions (ratio) method reverses s/g effects: larger raw volumes for all structures in males become slightly larger volumes for all structures in females after ICV ratio normalization. By contrast, when the same data were analyzed using ICV covariance, s/g differences in individual structure volumes all but disappeared (Voevodskaya et al., 2014) or became small and sparse (Jäncke et al., 2015; Pintzka et al., 2015; Sanchis-Segura et al., 2019).

Another issue is the choice of ICV versus TBV as the denominator or covariate for correcting regional volumes. With the emergence of automated segmentation methods, it is now much easier to measure TBV, which is the sum of all GM and WM. Compared to ICV, which remains relatively stable through adulthood, TBV declines some 16 % with aging, as sulci and ventricles expand and brain tissue steadily shrinks between the third and eighth decade (Whitwell et al., 2001). Thus, structures that are normalized to TBV provide a more direct estimate of volume relative to this changing baseline. In a meta-analysis involving 1,239 participants across eight research sites, Fjell et al. (2009) found that most individual brain structures were slightly larger in males after covarying for ICV, but few s/g differences survived correction by TBV covariance. Very similar findings were reported by Jäncke et al. (2015) in their analysis of 856 healthy participants: some structures were slightly larger in males following ICV covariance but there were fewer differences and they tended to be larger in females following TBV covariance. These authors reported that s/g accounts for 0.0–2.4% of the variance in volume of particular structures following ICV covariance (equivalent to $d < 0.31$) and 1.4 % or less of the population variance following TBV covariance (equivalent to $d < 0.24$). Finally, in their very large study, Ritchie et al. (2018) found that 85 % of the structures exhibiting s/g volume differences were eliminated with TBV covariance correction. The remaining 15 % of structures showed small effect sizes ($d < 0.25$), with some larger in females and others larger in males.

S/g differences in regional volumes are thus highly sensitive to the method used to correct for individual differences in global brain volume, which may contribute to the lack of reproducibility across studies (see Sections 3.1 and 3.2). Importantly, such correction does not apply to cortical thickness, a one-dimensional measure that does not vary with ICV (Potvin et al., 2017, 2018a; Barnes et al., 2010); studies of cortical thickness that attempt to correct for ICV thereby end up distorting estimates of s/g difference (see Section 3.3).

2.4. Effect of brain segmentation method on male-female comparisons

A last methodological concern involves the choice of brain segmentation software, which is critical for assessing regional differences in both volumes and cortical thickness. As datasets have grown larger, researchers have increasingly turned to automated methods for segmenting, or delineating, brain images into component structures. But such automation has a downfall, in that it requires every brain to be fitted (or, "registered") to a template, whether by stretching or shrinking in three dimensions (for volume-based methods such as voxel-based morphometry, VBM) or two dimensions (for surface-based methods such as Freesurfer and FIRST/FSL). Due to the inherent variability between individual brains, every automated fitting produces some distortion, which may affect male or female brains differently depending on the degree of registration and idiosyncrasies of the template used. Not

only are such templates unique to each analysis package, they are themselves constructed from a finite population of “normal” brains, often based on unequal numbers of males and females. The magnitude of this problem is further illustrated by the substantial differences between brain templates constructed from different ethnic populations (Yang et al., 2020b).

Such distortions have been demonstrated in various studies that have compared automated segmentation to manual tracing, still considered the “gold standard” for brain morphometry. Thus, Morey et al. (2009) found that two surface-based methods, FreeSurfer and FIRST/FSL, both inflate subcortical volumes by 4–9 %, with Freesurfer producing greater inflation of the amygdala and FIRST inflating the hippocampus to a greater degree. Grimm et al. (2015) compared VBM and Freesurfer to manual measurement of the same two structures and found that both programs deviated from manual tracing, with the accuracy of each method varying by structure. (VBM was more accurate for the amygdala, Freesurfer for the hippocampus.) More recently, Makowski et al. (2018) extended this comparison to the striatum, pallidum, and thalamus and found again that FSL and Freesurfer overestimate these volumes, with the error greater for smaller (e.g., pallidum) compared to larger structures (see also Perlaki et al., 2017; Schoemaker et al., 2016). By contrast, comparison of Freesurfer and manual volumetry on cortical structures (frontal gyri) found that Freesurfer produced systematically smaller volumes, with much smaller variance than manual methods, perhaps reflecting registration to a template that inadequately reflects the

diversity of human frontal cortex (Cox et al., 2016).

It is not known whether these distortions differentially affect the measurement of male versus female brain structures, but the fact that automated errors are greater for smaller structures suggests that they may. Indeed, in their meta-analyses of amygdala volume sex difference, Marwaha et al. (2017) found smaller male-female differences among studies using manual, as opposed to automated segmentation, whether the data were reported as raw volumes or as volumes normalized to ICV or TBV. Kennedy et al. (2009) similarly found discrepancies between VBM and manual measurements of cortical structures such that “sex differences all but disappeared” when manually-determined ICV was used as a covariate.

In sum, accurate comparison of specific brain structures between men and women depends on several methodological factors: structural segmentation software, normalization algorithm, and choice of normalization standard. Variations among these methods likely contribute to the wide range of findings and lack of reliable s/g effects in the neuroimaging data we discuss next.

3. Survey of structural brain sex/gender differences

The literal meaning of the term “dimorphic” is “two shapes.” Almost no studies have compared the actual shape of brain structures between males and females but hundreds have compared structure size, so we restricted this part of our survey to studies that report on the volume,

Table 1
Sex/gender difference in subcortical structure volumes.

First author (year)	N	Mean age (range)	Caudate	Putamen	Pallidum	Thalamus	Cerebellar WM	Cerebellar GM	Hippocampus	Amygdala	Correction method	Segmentation Method
Filipek (1994)	20	27.2	F	n.s.	n.s.	-	n.s.	n.s.	F	n.s.	TBV ratio	Manual
Caviness (1996)	30	9.2	F	n.s.	F	-	n.s.	n.s.	F	M	TBV ratio	Manual
Giedd (1997)	121	10.8	F	n.s.	M	-	-	-	n.s.	n.s.	TBV covariance	Manual
Mu (1999)	619	64.3	-	-	-	-	n.s.	n.s.	n.s.	n.s.	ICV ratio	Manual
Xu (2000)	331	55.2	-	-	-	F	F	F	-	-	ICV ratio	Manual
Goldstein (2001)	48	37.9	n.s.	n.s.	n.s.	-	-	-	n.s.	n.s.	ICV ratio	Manual
Good (2001)	465	32.1	-	-	-	-	M	M	M	M	VBM modulation	VBM
Gur (2002)	116	26	-	-	-	-	-	-	n.s.	n.s.	ICV covariance	Manual
Sowell (2002)	35	11.1	F	n.s.	n.s.	F	-	-	-	-	ICV ratio	Manual
Pell (2008)	176	33	-	-	-	n.s.	-	-	n.s.	-	ICV covariance	VBM
Neufang (2009)	46	11.3	F	F	F	-	-	-	F	M	ICV covariance	VBM
Fjell (2009)	1143	46.8	n.s.	M	M	M	n.s.	n.s.	M	M	ICV covariance	Freesurfer
Barnes (2010)	78	59.5	n.s.	n.s.	n.s.	-	-	-	n.s.	n.s.	TBV covariance	Freesurfer
Paus (2010)	579	15	F	M	M	M	-	-	n.s.	M	ICV covariance	Freesurfer
Goodro (2012)	184	50.8	n.s.	n.s.	n.s.	n.s.	-	-	n.s.	n.s.	ICV covariance	FSL FIRST
Brain Dev (2012)	325	10.9	n.s.	M	n.s.	n.s.	M	-	-	-	TBV covariance	ANIMAL
Rijpkema (2012)	463 (1.5T)	22.9	n.s.	M	M	-	-	-	-	-	TBV covariance	FSL FIRST
Inano (2013)	541 (3T)	22.7	n.s.	M	M	-	-	-	-	-	ICV covariance	Freesurfer
Inano (2013)	861	56.1	n.s.	F	F	F	F	F	F	F	ICV ratio	Freesurfer
Tang (2013)	40 (Beijing)	21.2	n.s.	M	M (right)	n.s.	-	-	n.s.	n.s.	TBV matched	FSL FIRST
	40 (Cambridge)	21	n.s.	M	n.s.	n.s.	-	-	n.s.	n.s.		
Ruirok (2014) <i>meta-analysis</i>	2186 16 studies	(10-47)	n.s.	M	n.s.	F	-	M	M	M	VBM modulation	VBM
Voevodskya (2014)	406 (PIVUS)	75	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	ICV covariance	Freesurfer
	223 (ANDI)	76	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.		
Jäncke (2015)	856	42.4	n.s.	n.s.	n.s.	F	n.s.	n.s.	n.s.	n.s.	ICV covariance	Freesurfer
			n.s.	n.s.	n.s.	F	n.s.	n.s.	n.s.	n.s.	TBV covariance	
Joel (2015)	855	27.6	F	-	-	F (left)	-	F	F	-	VBM modulation	VBM
	267	48.7	-	M	-	M	-	M	-	-	none	Freesurfer
Király (2016)	103	32	M (right)	n.s.	n.s.	M (left)	n.s.	F	F	n.s.	ICV	FSL FIRST
Pintzka (2015)	966	58.5	n.s.	n.s.	n.s.	n.s.	n.s.	M	n.s.	n.s.	ICV covariance	Freesurfer
	304	58.2	n.s.	n.s.	n.s.	n.s.	n.s.	M	n.s.	M	ICV matched	
Reardon (2016)	184	12.9	n.s. (striatum)	M	n.s.	-	-	-	-	-	TBV covariance	MAGeT
Tan (2016) <i>meta-analysis</i>	2183 29 studies	0-79	-	-	-	-	-	-	n.s.	-	various	various
Marwaha (2017) <i>meta-analysis</i>	1560 13 studies	(13-63)	-	-	-	-	-	-	-	n.s.	various	various
Potvin (2016/2018b)	2712	47.6	n.s.	M	M	M (right)	-	-	n.s.	M	ICV covariance	Freesurfer
Coupé (2017)	2944	39.7	F	n.s.	n.s.	F	-	-	n.s.	n.s.	ICV ratio	volBrain
Ritchie (2018)	5216	61.7	n.s.	M	M	n.s.	-	-	n.s.	M	TBV covar	FSL FIRST
Lotze (2019)	2838	52.4	n.s.	M	n.s.	M	-	-	M	M	VBM modulation	VBM
Sanchis-Segura (2019)	356	22.3	n.s.	M	M	n.s.	-	-	n.s.	M	ICV covariance	SPM12

Chronological listing of large or highly-cited MRI studies reporting mean volume of subcortical structures in healthy males and females, corrected for individual TBV or ICV. Pink cells report significantly larger volume in females (F). Blue cells report significantly larger volume in males (M). Yellow cells report no significant difference (n.s.). Unshaded cells were not reported. Abbreviations: ANDI (healthy controls from Alzheimer’s Disease Neuroimaging Initiative; PIVUS (Prospective Investigation of Vasculature in Uppsala Seniors Study).

Citations not included in article text: (Caviness et al., 1996; Goodro et al., 2012; Gur et al., 2002; Inano et al., 2013; Király et al., 2016; Mu et al., 1999; Neufang et al., 2009; Pell et al., 2008; Reardon et al., 2016; Tang et al., 2013; Xu et al., 2000).

thickness, or cross-sectional area of specific brain structures.

3.1. Negligible differences in subcortical structural volumes

The most oft-repeated claims about regional brain differences between men and women concern subcortical structures, which have been examined using MRI since the mid-1990s. Table 1 lists the results of 33 studies comparing the volume of subcortical structures between healthy human males and females, after normalizing for individual ICV or TBV. This is not a fully systematic list, but includes the most highly-cited studies, together with larger, more recent studies, some with well over 1000 participants. As is evident from Table 1, no subcortical structure has been found larger in males or females across all studies.

This table does not include the corpus callosum, which is considered subcortical and was the first human brain structure to be called “sexually-dimorphic.” Because of its relationship to lateralization, corpus callosum s/g effects are discussed in Section 3.4.

The next structures to be declared sexually-dimorphic were the hippocampus and amygdala. In a highly-cited review, Cahill (2006) depicted a mid-sagittal brain with many structures shown in pink (larger in women) and others shown in blue (larger in men). Although widely-reproduced (e.g., Purves et al., 2018 p. 568), this figure was based on just one small study in which most s/g differences were not statistically significant (Goldstein et al., 2001). Similar claims about the amygdala and hippocampus are oft-repeated (Cosgrove et al., 2007; Gur et al., 2010; Hines, 2011; Knickmeyer et al., 2014), citing Goldstein et al. (2001) or even earlier reports (Filipek et al., 1994; Giedd et al., 1996), and argued to be pertinent to s/g differences in memory (Andreano and Cahill, 2009), anxiety and depression (Giedd et al., 2012; Hammerslag and Gulley, 2016; Whittle et al., 2014). Claims about hippocampal and amygdala volume differences have also penetrated popular reporting, where they have been invoked to explain male-female differences in learning and emotion (Goldman, 2017; Gurian and Stevens, 2004; Murray, 2020).

However, these assertions ignore larger and more recent studies that have failed to confirm male/female differences in either hippocampal or amygdala volume (Table 1). The largest single study (Ritchie et al., 2018) found no difference in TBV-corrected hippocampal volume, whereas a meta-analysis of 29 studies found a mere 0.6 %, non-significant difference in hippocampal volume between males and females (Tan et al., 2016). According to Jäncke et al. (2015), s/g accounts for just 0.3 % ($d \approx 0.10$) of TBV-adjusted hippocampal volume.

The results are similar for the amygdala, where 16 of 29 analyses have found no significant volume difference between males and females (Table 1). Meta-analysis of 11 studies by Marwaha et al. (2017) found a slight, non-significantly larger amygdala volume in males (0.06 % right, 2.5 % left), and no difference was detected in the study of 2,944 participants by Coupé et al. (2017). In other large studies, the amygdala was one of the handful of “dimorphic” structures to survive ICV covariance in Potvin et al. (2016; 2018a) and TBV covariance in Ritchie et al. (2018), but the effect size is small ($0.08 \leq d \leq 0.18$; Fjell et al., 2009; Ritchie et al., 2018), amounting to 2% larger volume in males (Pintzka et al., 2015), or 1% of the total variance after TBV correction (Jäncke et al. (2015)). Thus, although often portrayed as settled science, s/g difference in amygdala volume is very modest, and its significance is highly sensitive to brain size correction method.

The caudate nucleus is another structure long described as sexually-dimorphic: proportionally larger in females and speculated to contribute to their lower risk for ADHD (Durston et al., 2001; Giedd et al., 2012). As with the hippocampus and amygdala, this claim is still referenced to a few early (Giedd et al., 1997) and very small studies (Filipek et al., 1994; Sowell et al., 2002), whereas larger and more recent studies have failed to confirm it (Table 1). It is possible the effect is confined to childhood and early adolescence, since three studies (Giedd et al., 1997; Sowell et al., 2002; Paus, 2010) report a larger ICV-corrected caudate volume in girls averaging 15 years of age. Nonetheless, the effect size is quite small

($d = 0.16$, per Paus, 2010) and was not confirmed in a comparably-sized study of children averaging 11 years of age (Brain Development Cooperative Group, 2012). Given the substantial data indicating an absence of caudate volume difference between adult men and women, any difference in adolescence could be a fleeting effect of girls’ earlier entry into puberty.

Collectively, the findings on other subcortical structures—including the pallidum, thalamus, cerebellum (Table 1) and nucleus accumbens—also fail to demonstrate consistent s/g differences in volume. If any finding is more reliable, it is a larger putamen in males, according to several larger studies (e.g., Lotze et al., 2019; Potvin et al., 2016, 2018b; Ritchie et al., 2018; Ruigrok et al., 2014). Once again, reported effect sizes are small ($0.04 \leq d \leq 0.25$ in studies by Fjell et al., 2009; Paus, 2010; Ritchie et al., 2018), corresponding to a TBV-corrected volume difference of 2.7 % in Rijpkema et al. (2012). However, no significant putamen difference was found in several other large studies: Coupé et al. (2017), Fjell et al. (2009), Jäncke et al. (2015), Pintzka et al. (2015) and Voevodskaya et al. (2014). Such variability was endorsed by Herting et al. (2018), who assessed male/female difference in subcortical volumes across three different international samples and found a striking lack of consistency between them.

In sum, many studies have now examined s/g differences in volume of the major subcortical brain structures, adjusted for overall brain size. Most claims of “sexual dimorphism” have not been reliably replicated and the two structures that have been found to differ in a plurality of studies—larger putamen and amygdala in males—do so by a mere 1–3 % volume difference.

3.2. Small effects and poor consensus for regional cortical volumes

Compared to subcortical structures, a coherent synthesis of s/g findings within the cerebral cortex is more difficult to achieve. Few papers report the volumes of every gyrus or major zone within the cerebral cortex. Rather, most studies analyze the entire cortical mantle but report only those regions that differ in volume between males and females above a certain threshold of statistical significance. Male/female similarity is not documented and differences are rarely reported as absolute or even proportional volume differences, making it difficult to compare quantitatively between studies.

Our overall sweep collected 25 large or highly-cited studies that have reported on regional volume s/g differences since the year 2000 (Suppl. Table 1). The average age of participants ranged from birth to late middle-age and the image analysis and brain size correction methods are also diverse. All of these factors could contribute to the wide range of results reported. Nonetheless, the simplest conclusion from this sizeable collection is that no particular gyri are reliably larger in males or females. In most studies, the majority of cortical volumes do not differ significantly (after correcting for overall brain size), and the differences reported in individual studies have been reproduced only sporadically in other studies.

In attempting to find patterns in this data, we sorted this set of 25 studies by several criteria, beginning with the method of brain/body size correction, which as discussed above can have dramatic impact on s/g findings. For instance, the one study that normalized cortical structure volumes by individuals’ height (Raz et al., 2004) resulted in all structures larger in males except for the inferior parietal lobule. But as discussed in Section 2.3, height is a poor covariate for normalizing brain volumes. Four of the 25 studies (Chen et al., 2007; Goldstein et al., 2001; Sowell et al., 2002; Voevodskaya et al., 2014) used a ratio method to normalize regional volumes to individuals’ TBV or ICV. As discussed above, this method is inferior to covariance correction and there is virtually no agreement among these four regarding cortical s/g differences. The majority of studies used either covariance or atlas-based methods to normalize for overall brain size, but findings within these subgroups are not well aligned, as discussed below. Finally, four studies compared cortical volumes between male and female brains that were

pre-matched for TBV or ICV (e.g., larger female compared to smaller male brains). This procedure eliminates the brain size normalization step entirely and is considered the gold standard for removing the influence of body size on brain s/g comparisons. However, as Table 2 shows, there is almost no overlap in findings between these four studies when taking laterality and subzones into account (Luders et al., 2009; Sanchis-Segura et al., 2019; Voevodskaya et al., 2014; Wang et al., 2012). Despite the strength of this approach, sample sizes are inevitably small and the participants varied in age considerably across studies.

If instead of attempting to reconcile all of the studies in Suppl. Table 1, we focus only on the largest (>300 participants), this eliminates many older studies and narrows down the methodologies for segmentation and brain size normalization (Table 3). Of course, larger studies also provide greater statistical power to resolve small differences and rule out false positives.

Three of the five most recent studies in Table 3 each included more than 2,700 participants. Of these, there is reasonable agreement between the two that used Freesurfer to perform cortical segmentation: studying 5,216 men and women and correcting for TBV covariance, Ritchie et al. (2018) found no s/g volume difference in 65 % of cortical regions, whereas 20 % of regions were larger in males and 15 % of regions were larger in females. Again, effect sizes for these s/g differences were small, with Cohen's d values ranging from 0.09 to 0.22. Among the regions found to differ by s/g, many were also detected by Potvin et al. (2017, 2018a) in their study of 2,713 participants that accounted for brain size using ICV covariance. These two studies agree in finding slightly larger volumes in males' fusiform, entorhinal, medial orbitofrontal, and lateral occipital gyri, together with the cingulate isthmus. Both studies also agree in finding larger volumes in females' superior parietal lobules and caudal medial frontal gyri. But they disagree with regard to the significance of s/g effects in the pars triangularis, lateral orbitofrontal, pre-central, post-central, paracentral, precuneus and lingual gyri, as well as in the laterality of volume differences in the medial frontal, pars opercularis, inferior parietal lobule, superior temporal, lateral occipital and cingulate isthmus.

The findings are notably quite different when we consider s/g differences based on VBM segmentation. The two largest such studies, a meta-analysis by Ruigrok et al. (2014) and an empirical study by Lotze et al. (2019), found some similarities: greater male volumes of the anterior parahippocampal gyrus; greater female volumes in select pre-frontal, inferior parietal, superior temporal and cingulate cortices, overlooking several laterality discrepancies. But these studies diverged regarding many other structures: frontal pole, superior parietal lobule, precuneus, insula, fusiform and several occipital gyri. Moreover, s/g findings were entirely different in three other analyses based on VBM or a related volume-based method, DARTEL (Joel et al., 2015;

Sanchis-Segura et al., 2019; Liu et al., 2020). Thus, even among some impressively large studies, the specific findings about cortical regions expanded in men or women do not replicate between them and especially, across studies using different MRI processing pipelines (Sanchis-Segura et al., 2019).

In sum, MRI studies over the past two decades have extensively analyzed, but failed to find evidence of universal s/g difference in the volumes of specific cortical regions. An obvious problem, particularly with early studies, is that they were underpowered to detect small differences (Button et al., 2013). However, even when we restrict our view to the largest and most recent studies, s/g differences are very small and markedly discrepant across populations and image analysis methods. They are not, in other words, species-wide s/g differences, nor do most findings align with predicted areas of cortical s/g difference based on well-described behavioral differences between men and women (see Section 8.1).

3.3. Cortical thickness: conflict between MRI and post-mortem measures

Along with cortical volumes, the thickness of cortical gray matter is often said to vary by s/g, with most reviews stating it is greater in females (Cosgrove et al., 2007; Sacher et al., 2013). Male/female difference in gray matter thickness has been proposed to contribute to gender differences in behavior and cognition (Luders et al., 2006a, b), including gender identity (Luders et al., 2012; Zubiaurre-Elorza et al., 2013). Moreover, focal changes in cortical thickness have been identified after brain training regimens such as motor learning (Sale et al., 2017), so it is reasonable to expect that differences in life experience between men and women will be reflected in this measure, at least in certain brain areas.

However, a synthesis of current findings does not show reproducible differences in cortical thickness between males and females, either globally or in specific cortical zones (Table 4). The earliest measures of cortical thickness were taken in postmortem histological studies, none of which detected significant male/female differences. Due to the painstaking nature of such measurement, such studies tend to be small. But even the largest one (94 brains; Pakkenberg and Gundersen, 1997) found only a 4% average difference, in this case, thicker in men (2.72 mm vs. 2.61 mm; $p = 0.064$, $d = 0.41$).

With the advent of MRI measurement, more explicit claims of s/g difference in cortical thickness have emerged. We found 21 highly-cited or recent MRI studies that have addressed the question, about evenly divided between finding no difference or greater thickness in females (Table 4). Only four MRI studies report significantly greater cortical thickness in males, though not at all ages (e.g., Salat et al., 2004). Raznahan et al. (2010) reported thicker gray matter in most cortical areas in adolescent boys. However, other studies of adolescents report

Table 2
Cortical volume sex/gender differences in samples matched for brain size.

First author (year)	# M/F	Mean Age	Larger in males	Larger in females	No difference	Matched for:
Luders (2009)	24/24	43.4	none	L. orbitofrontal, superior temporal, superior frontal	Most other regions	TBV
Wang (2012)	35/35	21.2	L. supplementary motor; R. middle frontal, middle temporal, lingual, supramarginal	Bilateral medial orbitofrontal, parahippocampal; L. middle frontal; R. precentral	Majority of cortex, especially parietal, temporal & occipital	TBV
Voevodskaya (2014)	21/21	75.0	none	none	All cortical areas	ICV
Sanchis-Segura (2019)	74/74	21.9	L. parahippocampal; R. lingual & fusiform gyri	Bilat. superior orbitofrontal; R. middle occipital	Majority of cortex, esp. frontal, temporal & superior parietal	ICV

Results of studies that compared regional cortical gyrus volumes between groups of male and female participants matched for overall brain size, either total brain volume (TBV) or intracranial volume (ICV). Color coding as in prior tables. Abbreviations as above plus: Bilat. (bilateral), L. (left), R. (right).

Table 3

Sex/gender difference in regional cortical volumes.

First author (year)	N	Avg. age	FRONTAL			SENSORY-MOTOR			PARITAL			TEMPORAL			OCCIPITAL			LIMBIC		Segment. Method					
			SFG	MFG	IFG	Orbito- frontal	Pre- central	Post- central	Para- central	Sup. lobule	Inf. lobule	Pre- cuneus	STG	MTG	ITG	Parahippo- campal	Fusi- form	Cuneus	Lingual	Super- ior	Lateral	Cingulate	Insula		
Good (2001)	465	32.1				R. lateral orbital					R.		L. anterior superior								R.		VBM		
Chen (2007)	411	46.7											R. transverse	R.								Bilat.		VBM	
Brain Dev. Coop (2012)	325	10.9													R.	L. ant.							ANIMAL		
Ruijgrok (2014) meta-analysis	2186	9.5- 46.7				R.	L.		R.			R. oper- culum		R. planum temporale & Heschl's			Bilat. ant.					Bilat. post., L. ant.	R.	VBM	
						Bilat. caud.										L. post.						Bilat. Post.		Freesurfer ICV ratio	
Voevodskaya (2014)	406	PIVUS																						Freesurfer ICV covar.	
Joel (2015)	855	27.6				Bilat. g. rectus; L. orb., triang.				Bilat.													VBM		
Knickmeyer (2017)	756	0.1											Bilat.			Bilat. posterior	L. anterior	L.					Bilat. post.		EMR
Potvin (2017, 2018a)	2713	47.6			R. caud.	Bilat. triang., R. orb.	Bilat. medial	Bilat.			Bilat.	R.	R.	L.			Bilat. entorhinal	Bilat.		Bilat.		Bilat.	Bilat. isthmus	R.	Freesurfer
Ritchie (2018)	5216	61.7			Bilat. caud.	Bilat. opercu- laris	Bilat. medial; R. lateral		Bilat.	Bilat.	Bilat.	L.		R-STG; L. transverse			Bilat. entorhinal	Bilat.				R.	L. ant.	R.	Freesurfer
Lotze (2019)	2838	52.4			Bilat. front. pole	Bilat. BA 46	Bilat. triang.	Bilat. medial	Bilat. premot.		Bilat.	Bilat.	Bilat.	Bilat.			Bilat.	Bilat.	Bilat.	Bilat.	Bilat.	Bilat.	Bilat. anterior		VBM
Sanchis-Segura (2019)	356	22.0						Bilat. sup. orbital; R. med. orbital								R.								DARTEL	
Liu (2020)	976	28.2			Bilat.	Bilat.	Bilat.	Bilat. medial	Bilat.	Bilat.	Bilat.	Bilat.	Bilat.		L.	L.	Bilat.	Bilat.	R.	R.	Bilat.	Bilat.	Bilat.	DARTEL	

Male/female group differences reported in studies measuring regional cortical volumes. This table includes only studies with more than 300 participants, and all volumes are normalized for individual TBV or ICV. Cell shading as in previous tables. For a larger collection of 25 large and highly-cited studies, see Supplemental Table 1. Abbreviations as above and: PIVUS (Prospective Investigation of Vasculature in Uppsala Seniors study); BA (Brodmann's area); IFG (inferior frontal gyrus); ITG (inferior temporal gyrus); MFG (middle frontal gyrus); MTG (middle temporal gyrus); SFG (superior frontal gyrus); STG (superior temporal gyrus); ant. (anterior); caud. (caudal); front. (frontal); med. (medial); mid. (middle); post. (posterior); orb. (IFG, pars orbitalis); triang. (IFG, pars triangularis).

no difference (Gennatas et al., 2017; Mills et al., 2014) or greater thickness in girls (Paus et al., 2017). The only study to address the issue in infants (0–2 years) found no difference in cortical thickness or in the structural and functional networks that correlate with it (Geng et al., 2017). Finally, a recent study reported greater thickness in male adults within two different populations (Chinese and American), but the precise regions differing between men and women showed no agreement across populations (Yang et al., 2020a).

By contrast, 10 independent studies, including those based on the largest samples, report predominantly thicker cerebral cortex in females (Table 4), albeit the difference is small ($d \leq 0.26$). According to Potvin et al. (2017), s/g accounts for less than 2% of variance in cortical thickness. S/g differences were more pronounced in parietal and frontal gyri, as compared to temporal and occipital zones according to some (Ritchie et al., 2018; Sowell et al., 2007; Zubiaurre-Elorza et al., 2013), but not all (Govindarajan et al., 2014; Potvin et al., 2017) studies.

Unlike measures of regional brain volumes or surface area, cortical thickness is largely insensitive to TBV or ICV (Potvin et al., 2017, 2018a; Barnes et al., 2010; see Table 4). According to Im et al. (2008), one consequence of this lack of scaling is that studies that do normalize cortical thickness by ICV or TBV end up over-correcting s/g difference, leading to an exaggeration of female advantage.

Taken together, existing data do not demonstrate a reliable s/g difference in cortical thickness. Although several large studies found it is greater in females, all were performed using Freesurfer, whereas other large MRI studies that used different segmentation methods found either no difference (Gennatas et al., 2017) or greater thickness in males (Raznahan et al., 2010). Freesurfer accuracy is further questioned by Govindarajan et al. (2014), who found significant s/g differences in scattered cortical regions measured in a 1.5 T scanner, but no difference in a separate cohort scanned with the higher resolution afforded by a 3 T magnet. (See also, Martínez et al., 2015; Mitricheva et al., 2019; Redolfi et al., 2015.) Finally, the fact that MRI analyses lean toward greater

thickness in females whereas post-mortem histology (e.g., Pakkenberg and Gundersen, 1997) finds slightly greater thickness in males should give us pause, since the latter has higher resolution and is arguably the gold standard for assessing tissue thickness.

3.4. Interhemispheric connections: corpus callosum, anterior commissure, interthalamic adhesion

Before all this focus on cortical structure, the most widely disseminated claim about human brain s/g difference was that women have a larger corpus callosum than men. Together with the anterior commissure and interthalamic adhesion, corpus callosum morphology is frequently invoked to argue that women have stronger interhemispheric connections and thereby, less lateralized cerebral function than men. According to one influential hypothesis (Witelson and Nowakowski, 1991), prenatal testosterone enhances pruning of callosal axons, such that males end up with fewer interhemispheric connections and greater structural and functional lateralization. To evaluate this theory, we first consider the anatomical evidence for male/female difference in the area or volume of the corpus callosum, anterior commissure and interthalamic adhesion; s/g difference in functional lateralization will be addressed in the following section.

As the largest of the inter-hemispheric connections, the corpus callosum has been the focus of most of this research. The original claim that women have a larger corpus callosum was based on a mere 14 brains, but nonetheless published in *Science* (DeLacoste-Utamsing and Holloway, 1982), attracting considerable public interest that remains lodged in popular writing today (Eliot, 2011; Jensen and Nutt, 2015; Goldman, 2017; Methil, 2019). However, the actual findings were controversial from the start. DeLacoste-Utamsing and Holloway (1982) did not actually find a difference in total corpus callosum cross-sectional area, but only a marginal difference in the splenium ($p = 0.08$). Although the same authors did extend their finding to total callosal area (De Lacoste

Table 4

Sex/gender difference in cortical thickness.

First Author (Year)	Measure	# M/F	Mean Age	Greater cortical thickness in:	Correction Method	Segmentation Method
Witelson (1995)	Histology	4/5	50.0	n.s. (M= 2.9mm, F=2.7 mm)	None	Manual
Pakkenberg & Gundersen (1997)	Histology	62/32	56.1	n.s. (M=2.7 mm, F=2.6 mm)	None	Manual
Rabinowicz (2002)	Histology	6/5	16.8	n.s. (M=2.6 mm; F=2.7 mm)	None	Manual
Salat (2004)	MRI	13/18	22.8	n.s.	ICV covariance	Fischl & Dale (2000)
		7/10	48.6	M	ICV covariance	
		16/42	76.6	n.s.	ICV covariance	
Im (2006)	MRI	31/21	26.1	n.s.	None	CLASP
Luders (2006)	MRI	30/30	24.9	F	Atlas Scaled	CLASP
				F	Atlas Scaled	Custom
Sowell (2007)	MRI	90/86	32.4	F (Bilat. posterior temporal; R. inferior parietal; L. ventral frontal)	None	Custom
				F (R. inferior parietal, posterior temporal)	TBV Matched	
Im (2008)	MRI	83/65	25.0	n.s.	ICV Scaling	CLASP
Lv (2010)	MRI	90/94	38.4	F	None	CLASP
Raznahan (2010)	MRI	153/131	14.1	M	None	CIVET(CLASP)
Crespo-Facorro (2011)	MRI	45/31	28.1	n.s.	None	BRAINS2
Zubiaurre-Elorza (2013)	MRI	29/23 (cis)	30.1	F (select frontal & parietal areas)	None	Freesurfer
Creze (2014)	MRI	16/12	25.5	M (R. superior parietal & postcentral)	ICV covariance	Freesurfer
				F (ant. cingulate and superior frontal)		
Govindarajan (2014)	1.5T MRI	74/77	46.9	F (Bilat. precentral; L. sensorimotor; R. transverse temporal, isthmus, cingulate, cuneus, lingual)	None	Freesurfer
				n.s.		
Mills (2014)	MRI	124/164	14.9	n.s.	None	Freesurfer
McKay (2014)	MRI	327/511	43.0	F (very slight difference)	None	Freesurfer
Wierenga (2014)	MRI	92/43	12.6	n.s.	None	Freesurfer
Vijayakumar (2016)	MRI	49/41	15.9	n.s.	None	Freesurfer
Paus (2017)	MRI	479/509	15.0	F	None	Freesurfer
Geng (2017)	MRI	61/57	32 wks	n.s.	None	Freesurfer
Gennatas (2017)	MRI	541/648	15.5*	n.s. (1-2% larger in males)	None	DiReCT
Potvin (2017/2018a)	MRI	1362/1351	47.6	F (negligible)	ICV covariance	Freesurfer
Ritchie (2018)	MRI	2466/2750	61.7	F (sensorimotor, dlPFC, parietal)	TBV covariance	Freesurfer
Yang (2020a)	MRI	129/121 China	21.5	M (bilat. frontal, L. temporal)	ICV covariance	Freesurfer
				M (R. insula, L. cingulate isthmus)		

Studies reporting sex/gender difference in cortical thickness. Cell color coding and abbreviations as in previous tables. Age is in years, unless indicated. Abbreviations as above plus: cis (cis-gender subjects); dlPFC (dorsolateral prefrontal cortex).

* = Age range = 8–23; mean not reported.

Citations not included in article text: (Crespo-Facorro et al., 2011; Creze et al., 2014; Im et al., 2006; McKay et al., 2014; Rabinowicz et al., 2002; Vijayakumar et al., 2016; Witelson et al., 1995).

et al., 1986; Holloway et al., 1993), other neuroanatomists reported null or opposite effects (Suppl. Table 2A). By the mid-1990s, enough studies had been published to conduct meta-analyses (Table 5A), and two such reviews found that raw corpus callosum area is modestly larger in *males* ($0.21 < d < 0.27$), but disagreed about whether callosal area is the same (Bishop and Wahlsten, 1997) or larger in females (Driesen and Raz, 1995) when normalized to total brain size. This prompted a third meta-analysis by Smith (2005), who found callosal area to be modestly larger in females ($d=0.22$) when expressed as a ratio of mid-sagittal brain area and much larger ($d=0.62$), when expressed as a ratio of brain weight.

As MRI came to supplant post-mortem studies, sample sizes grew larger and brain size correction became easier. The fact that corpus callosum area and volume are sensitive to overall brain size led several groups to formally assess this relationship across sexes. Jäncke et al. (1997) were the first to show that larger brains have proportionally smaller corpus calloso irrespective of s/g, a result replicated by Leonard et al. (2008) and hypothesized to be due to the relative inefficiency of callosal transfer in bigger brains (Ringo et al., 1994). This was confirmed by Luders et al. (2014) who found that the s/g difference in corpus callosum thickness disappeared when male and female brains were selected to match in intracranial volume. However, two other analyses of males and females matched for ICV (Table 5B) came to diametrically

opposite conclusions: Sullivan et al. (2001) reported larger corpus callosum area in males, whereas Ardekani et al. (2013) and Shiino et al. (2017), studying the same sample, reported larger callosal area in females.

It is likely that much of this discrepancy is due to small samples. Few of the aforementioned studies (or those included in the meta-analyses) included more than 100 participants per group, making them grossly underpowered to detect an effect size of 0.22 (which requires $N \approx 700$). Focusing on the six largest studies (Table 5C), five have found no sex difference in corpus callosum area or volume (Good et al., 2001; Herron et al., 2012; Jäncke et al., 2015; Lenroot et al., 2007; Voevodskaya et al., 2014). Most recently, Potvin et al. (2016, 2018b) published normative MRI results from a very large sample (2,790 participants) which detected a significant, but only slightly larger corpus callosum volume in females, with sex accounting for less than 1% of the variance in that population ($d=0.17$).

Diffusion tensor imaging (DTI) is now permitting a more refined analysis of the corpus callosum and other white matter tracts. In a recent study of 433 young adults, Björnholm et al. (2017) found no significant difference in mean diffusivity (MD; a measure of white matter integrity) of the corpus callosum and small, spotty differences in fractional anisotropy (FA; a measure that integrates fiber density and myelination) but a significantly higher magnetization transfer ratio (MTR) in males;

Table 5

Sex/gender difference in corpus callosum (CC) size.

Part A: Meta-analyses

Meta-analysis	Measurement	Source	# Studies	Larger in:	d-value	Correction Method
Driesen & Raz (1995)	Absolute area of CC and splenium	Post-mortem & MRI	36	M	0.27	None
			11	F	-0.26	Ratio to brain weight or total midsagittal area
Bishop & Wahlsken (1997)	CC area	Post-mortem & MRI	41	M	0.21	None
Smith (2005)	Midsagittal area	Post-mortem & MRI	14	F	-0.22	Ratio to total midsagittal brain area
			8	F	-0.62	Ratio to brain weight

Part B: Corpus callosum sex/gender difference in samples matched for ICV (all MRI)

ICV-matched study	Measurement	#M/F	Mean age	Larger in:	Segmentation Method
Sullivan et al. (2001)	Midsagittal area	27/22	46.9*	M	Manual
Luders et al. (2014)	Thickness	24/24	43.4	none	Custom
Ardekani et al. (2013); Shiino et al. (2017)†	Midsagittal area; Volume	37/37	45.1*	F; F (genu)	Custom; VBM (DARTEL)

* Mean age was not reported for matched samples in these studies; value in table is the mean age of the parent sample.

† The same sample was used for both of these studies.

Part C: Corpus callosum sex/gender difference in largest MRI studies

Large studies	Measurement	# M/F	Mean Age	Larger in:	Correction Method	Segmentation method
Good et al. (2001)	Volume	265/200	32.1	none	None	VBM
Lenroot et al. (2007)	Midsag. area	209/178	13.2	none	None	ANIMAL
Herron et al. (2012)	Midsag. area	369/495	24.2	none	none	Custom VBM
Voevodskaya et al. (2014)	Volume	213/193	75.0	none	ICV covariance	Freesurfer
Jäncke et al (2015)	Volume	404/452	42.4	none	ICV covariance	Freesurfer
Potvin et al. (2016/2018b)	volume	1390/1400	47.6	F	ICV regression	Freesurfer

Best evidence on sex/gender difference in corpus callosum morphometry. A. Meta-analyses. B. Studies employing M/F samples matched for ICV. C. Sex/gender effects in the 6 largest studies. Cell color coding as in previous tables. d-values are positive for M > F, negative for F > M. See Supplemental Table 2 for a more complete collection of large and highly-cited studies of adults, plus studies assessing CC sex/gender difference in fetuses and children.

MTR is thought to largely reflect myelin content. However, this finding is opposite the results of a study of 408 adolescents by the very same research group, who reported higher MTR in females (Perrin et al., 2009). Moreover, a systematic review of DTI measures of the corpus callosum in schizophrenia found no effect of sex in either controls or patients in the majority of studies (Shahab et al., 2018). In sum, DTI measures do not support the claim that the corpus callosum is larger or more efficient in females.

A last issue is whether a s/g difference in the corpus callosum is present in early life, addressing the hypothesis that prenatal testosterone triggers pruning of interhemispheric fibers (Witelson and Nowakowski, 1991). Over the past three decades many studies have assessed male/female difference in the corpus callosum during fetal and childhood development, most using sample sizes too small to detect the modest effect sizes cited above. Nonetheless, the fact that 9 out of 13 studies of fetuses and children (Suppl. Table 2B) find no significant male/female difference, and the two largest of these studies (Achiron et al., 2001; Westerhausen et al., 2016) reported opposite findings, does not support this claim. As direct evidence against the pruning hypothesis, one study actually measured fetal amniotic testosterone levels and found no correlation with corpus callosum area in boys at ages 8–11 (Chura et al., 2010).

After the corpus callosum, the anterior commissure (AC) is the next largest connection between left and right cerebral hemispheres

(Table 6A). Like the corpus callosum, the AC has long been claimed to differ between males and females, bolstering the argument that females have stronger inter-hemispheric connectivity. However, the first study of AC cross-sectional area actually found it was larger in males (Demeter et al., 1988). This was followed by two studies by Allen and Gorski (1991, 1992) reporting larger AC area in females. Then came two studies finding no s/g difference (Highley et al., 1999; Lasco et al., 2002). In 2011, Choi et al. published the first MRI analysis of AC volume found no s/g difference in young adults but a larger AC volume in middle-aged women than men (after ICV covariance). Taken together, these six studies provide weak support, at best, for the claim of s/g difference in AC area or volume.

By contrast, one inter-hemispheric structure that does differ reliably between males and females is the massa intermedia, also known as the interthalamic adhesion (ITA; Table 6B), a small bridge of glia connecting the left and right thalamus that is entirely absent in 2–22% of healthy brains (Trzesniak et al., 2011). In their first analysis of the AC, Allen and Gorski (1991) also found that the ITA is larger and more often present in the brains of women than men, noting similar findings dating back to the 1940s. This has been confirmed by more recent studies, most of which have found that the ITA is more often present and/or is larger in women. However, in the largest and most recent report, Damle et al. (2017) used 3 T MRI and found no difference in ITA absence (both were only 4%) between males and females, albeit a 14 % larger ITA area in women. This

Table 6

Sex/gender difference in the anterior commissure and massa intermedia.

Part A: ANTERIOR COMMISSURE						
Study	Measurement	Method	# M/F	Mean Age	Larger in:	Brain Size Correction
Demeter et al. (1988)	Midsagittal area	post-mortem	22/11	60	M	none
Allen & Gorski (1991)	Midsagittal area	post-mortem	50/50	58.5	F	none
					F	none
Allen & Gorski (1992)	Midsagittal area	post-mortem	30/30	39.2	F	brain wt. ratio
Highley et al. (1999)	Midsagittal area	post-mortem	15/14	70.4	none	none
					none	none
Lasco et al. (2002)	Midsagittal area	post-mortem	57/43	50.7	none	brain wt. covariance
			46/47	23	none	
Choi et al. (2011)	Volume	3T MRI	36/51	47.5	F	ICV Covariance

Part B: MASSA INTERMEDIA						
Study	Measurement	Method	# M/F	Mean Age	Larger in:	ITA more common in:
Allen & Gorski (1991)	area	post-mortem	50/50	58.5	F	F
Snyder et al. (1998)	A-P length	MRI	30/22	27.7	n.r.	none
Nopoulos et al. (2001)	A-P length	MRI	53/59	28.5	n.r.	F
Erbagci et al. (2002)	A-P length	MRI	11/18	28.6	n.r.	none
Sen et al. (2005)	area	1.5T MRI	61/100	n.r.	F	n.r.
de Souza Crippa et al. (2006)	A-P length	1.5T MRI	26/12	29.7	n.r.	none
Ceyhan et al. (2008)	area	1.5T MRI	39/50	36.6	none	F
Takahashi et al. (2008a)	A-P length	MRI	46/35	24.5	F	F
Takahashi et al. (2008b)	A-P length	MRI	55/32	26.9	F	F
Damle et al. (2017)	A-P length	3T MRI	129/104	30.4	F	none

Sex/gender difference in size of the: A) the anterior commissure; B) massa intermedia, also known as the interthalamic adhesion (ITA). Other abbreviations: wt. (weight); n.r. (not reported); A-P (anterior-posterior).

Citations not included in article text: (Ceyhan et al., 2008; de Souza Crippa et al., 2006; Erbagci et al., 2002; Sen et al., 2005; Snyder et al., 1998; Takahashi et al., 2008a, b).

finding thus appears reliable, although little research has addressed the possibility that ITA presence and size are affected by overall brain size. That is, if larger brains are less able to mechanically support this small bridge across the third ventricle, the size and presence of ITA may be a function of brain geometry, not s/g per se. Only one study has examined this issue (Nopoulos et al., 2001) and found no relationship between ITA presence and TBV, but further examination is warranted. Of note, ITA absence has been associated with a greater incidence of schizophrenic disorders in both men and women, perhaps reflecting global brain atrophy (Trzesniak et al., 2011).

In sum, evidence across these three inter-hemispheric structures provides minimal support for the claim that the left and right hemispheres are anatomically more connected in women. Neither the corpus callosum nor anterior commissure is reliably larger in females. The ITA is, but as a small glial structure, it unlikely contributes to functional interhemispheric connectivity.

4. Sex/gender difference in brain lateralization and asymmetry

This leads us to the most prominent “sexual dimorphism” claimed for the human brain: that men’s brains operate in a meaningfully more lateralized or asymmetric manner than women’s. As we’ve seen, anatomical evidence from the corpus callosum and anterior commissure does not support sparser interhemispheric connectivity in males. But there is much other behavioral and brain activity research that has addressed this hypothesis. As we’ll see, the evidence using both measures indicates that s/g difference in brain lateralization is modest, at best.

The idea that cerebral lateralization differs between men and women was propelled by two seminal papers in the early 1980s. The first was a survey of sex findings on lateralization by McGlone (1980) and the second was a wide-ranging theory of cerebral lateralization introduced by Geschwind and Behan (1982) and expanded upon in three lengthy

papers (Geschwind and Galaburda, 1985a, b, c). McGlone (1980) relied heavily on clinical reports to argue that women are less lateralized for language and (to a lesser extent) spatial tasks than men; according to her review, left hemisphere lesions were more likely to result in aphasia when they occurred in males, compared to females. Geschwind and colleagues were focused on the mechanism of lateralization itself, which they hypothesized to involve prenatal testosterone and to be *less* pronounced in males. Both theories have been highly cited over the past 40 years.

But despite their wide reach, neither theory has held up to critical analysis. Clinical research—based on large studies of the incidence, type, and recovery from aphasia following left hemisphere damage—has not supported the claim that women are less left-dominant for language than men (Plowman et al., 2012; Watila and Balarabe, 2015). Nor do fMRI studies of language lateralization demonstrate a reliable s/g difference. Although a small, early study continues to be widely cited as finding greater male lateralization during one language task (Shaywitz et al., 1995), a subsequent meta-analysis (Sommer et al., 2008) of 26 studies found no s/g difference ($d = 0.01$) in the degree of lateralized activation during language processing. (See also Section 7.1.) Sommer et al. (2008) also found no evidence ($d = -0.11$, $p = 0.68$, $\kappa = 13$) for s/g difference in anatomical asymmetry of the planum temporale, a key structure for receptive language (i.e., Wernicke’s area). On the other hand, a more recent study of 2,337 young adults did detect significantly lower ($d = -0.22$) planum temporale asymmetry in women (Guadalupe et al., 2015), as would be predicted if language were modestly less lateralized in females.

Even more data on s/g laterality effects is available from a long history of research comparing the processing of simultaneous stimuli to left and right sides of the body. McGlone (1980) included such dichotic listening and visual tachistoscopic evidence to bolster the claim of greater lateralization in males. However, Hiscock and colleagues carried out four comprehensive reviews and found the preponderance of studies

detected no significant difference between males' and females' laterality for auditory (Hiscock et al., 1994) visual (Hiscock et al., 1995), tactile (Hiscock et al., 1999), or dual-task (Hiscock et al., 2001) processing. Overall, Hiscock and colleagues concluded that the s/g effect is a weak, population-level difference, accounting for just 1–2 % of the total variance in laterality. Voyer (1996) came to a similar conclusion based on meta-analysis of 396 comparisons in 266 studies, which calculated a pooled s/g difference in laterality of about $d = 0.08$, accounting for just 0.1 % of the population variance. A comparably small effect size ($d = 0.09$) was reported for the sex difference in dichotic listening according to meta-analysis by Sommer et al. (2008).

Importantly, Voyer (1996) also analyzed for publication bias and concluded that even this tiny, 0.08 effect size is likely an over-estimate based on the apparent under-publication of null effects. Ignoring publication bias, it would take more than 2200 participants to detect such a small, population-level s/g difference at the 95 % confidence level. Boles (2005) later put this to the test in a study of lateralized processing among 836 college-age students; as expected, most tasks failed to result in significant male/female laterality difference. For the 2 out of 6 tasks that did reach one-tailed significance, s/g accounted for less than 1% of the population variance. Like Voyer (1996) and Boles (2005) concluded that the s/g difference in hemispheric lateralization is extremely small, detectable only in extremely large samples, and unlikely to be meaningful in any clinical or other real-world context.

Another challenge to the theory of greater male lateralization is the fact that left-handedness—generally regarded as a sign of reduced cerebral dominance—is actually more common in males. Indeed, this was the one positive finding in Sommer et al. (2008): based on nearly a quarter-million participants, non-right handedness is 25 % more prevalent in males. This difference is actually what inspired Geschwind and Galaburda (1985a, b, c) to propose a model that is essentially the opposite of McGlone (1980): that males are *less* likely to develop the species-typical, left-hemisphere dominance, leading to a greater propensity for both left-handedness and various language-related disorders (e.g., dyslexia, stuttering, and autism). Specifically, they proposed that prenatal testosterone slows the growth of the left hemisphere such that males are relatively more right-hemisphere dominant than females, leading to more left-handedness, poorer language skill, and stronger spatial skill. But here too, in-depth reviews and meta-analyses found a lack of solid evidence for this theory (Berenbaum and Denburg, 1995; Bryden et al., 1994) or the hypothesized influence of prenatal testosterone on functional lateralization (Mercure et al., 2009; Pfannkuche et al., 2009).

More recently, the issue of s/g difference in lateralization has been addressed using resting fMRI connectivity. One of the first such studies, by Liu et al. (2009), found a greater degree of laterality among 131 male, compared to 169 female participants, although both groups exhibited the same left dominant activity in language and midline structures versus right-dominant activity in visual, parietal and insular cortices. Using a different, voxel-based activity measure on a sample of 913 young adults, Tomasi and Volkow (2012b) also found hemispheric dominance patterns to be largely similar between males and females even though their paper emphasized small islands of laterality differences within occipital, temporal, and frontal cortices. In a study of 603 participants, Agcaoglu et al. (2015) analyzed the lateralization patterns of different resting state networks and found they were largely similar between males and females, except for two minor differences that did not survive correction for false discovery. Finally, in the largest analysis of this type, Nielsen et al. (2013) used resting state functional connectivity to examine left- versus right-brain dominance in a sample of 1011 young adults; here again, distinct networks show different patterns of lateralization (e.g., left-lateralized language networks and right-lateralized attention networks), but there was no difference between males and females in these patterns.

It is notable that two of these studies of resting network lateralization (Nielsen et al., 2013; Tomasi and Volkow, 2012b) used data from the

same, 1000 Functional Connectomes Project, demonstrating how different methods of connectivity analysis can create or eliminate small s/g difference depending on the processing pipeline. Thus, the s/g findings of connection laterality are not robust across analysis methods, even using nearly identical participant samples. A similar conclusion was reached by Hausmann (2017).

Taken together, these various measures of cerebral lateralization—including clinical, behavioral, anatomical, and resting activity measures—demonstrate slight, if any, difference between men and women in the degree of functional hemispheric asymmetry (Hirnstein et al., 2019). Although s/g effects on lateralized brain function have been extensively analyzed over many decades, the collective data do not support the widespread belief that males' brains are meaningfully more lateralized than females'.

5. Sex/gender differences in neural connectivity

In contrast to the decades-long history of research on brain lateralization, attempts to map the influence of s/g on the overall connectivity plan of the brain, or “connectome,” are quite recent. Thanks to large, open-access MRI datasets, many papers have been published in the last few years comparing connectivity patterns in male and female brains. Such analyses are often focused on aging and brain disease, but here we address only the larger and more highly-cited studies that supply normative values for healthy males and females.

Approaches for comparing the connectivity patterns of male and female brains fall into two types: structural and functional. Structural connectivity has mostly been inferred from DTI tractography, which translates the three-dimensional diffusion of water into the density and direction of white matter pathways. Other structural connectome methods rely on complex statistical correlations between the volumes, thickness, or gray matter density of different cortical areas. Functional connectomes are constructed from fMRI measures of cerebral blood flow and correlational analyses of “resting” (untasked) activity in different brain areas. In addition to the wide range of data used, connectome studies are quite varied in their analytic tools and the measures they report to describe global connectivity patterns, as well as methods used to adjust for brain size. These variances, along with the absence of a universal definition of “connectome” leave us with a widely disparate set of findings and absence of consensus about s/g effects.

5.1. Sex/gender effects on the structural connectome

Altogether we identified 31 large or highly-cited studies that assessed sex/gender difference in the human connectome, 11 of which analyzed the structural connectome in adults (Suppl. Table 3A). Six of these 11 reported either no male/female differences (Hänggi et al., 2014; Mechelli et al., 2005; Zhao et al., 2015) or minor local differences in the face of global similarity in modularity (degree of network segregation into topological clusters) and “small-world” topology (Lim et al., 2015; Lv et al., 2010; Sun et al., 2015), meaning an organization dominated by local nodes with high levels of nearby connections. In addition, three of the studies reported higher average efficiency of networks in female brains (Duarte-Carvajalino et al., 2012; Gong et al., 2009; Yan et al., 2011), as emphasized in a review by Gong et al. (2011) but notably not replicated in two later studies (Dennis et al., 2013; Zhao et al., 2015).

Despite these mixed and modest findings, one structural connectome analysis received broad coverage and has been widely accepted as the definitive s/g difference in brain connectivity. The study, by Ingalkhalikar et al. (2014), did not report on network efficiency but instead focused on the contrast between intra-hemispheric connectivity, which was greater in males, versus interhemispheric connectivity, which was greater in females. This contrast was not seen earlier by Dennis et al. (2013) or Duarte-Carvajalino et al. (2012), but Ingalkhalikar et al. (2014) attributed the difference to their roughly two times larger participant

sample than earlier studies.

Indeed, most of the s/g effect sizes in Ingalkhalikar et al. (2014) were small, suggesting a very large sample was needed to detect them. However, subsequent analysis by Hänggi et al. (2014) was able to reproduce the intra- versus inter-hemispheric s/g contrast with just 138 participants. The caveat is that Hänggi et al. (2014) demonstrated that this contrast is a function of brain size, not s/g per se. Thus, they showed that larger brains have relatively lower inter-hemispheric connectivity and proportionally higher intra-hemispheric connectivity than smaller brains, regardless of s/g. This finding was confirmed by Martinez et al. (2017) and agrees with earlier findings on the corpus callosum (Leonard et al., 2008). DTI tractography is highly sensitive to the physical distance between nodes, which is greater in subjects with larger brains (males, on average). Since Ingalkhalikar et al. (2014) did not correct for brain size, whereas the studies by Duarte-Carvalhalino et al. (2012) and Dennis et al. (2013)—and indeed, all other studies in Suppl. Table 3A—did, the discrepancies appear to be explained by brain size rather than sample size or s/g per se.

In summary, estimates of brain connectivity based on structural MRI and DTI measures have not converged on a reliable pattern of s/g difference. Several studies report greater local (but not global) efficiency in females' brain networks, but the specific locales of such connections do not agree across studies. Larger brains appear to exhibit relatively more intra- than inter-hemispheric connectivity, but this is not a function of s/g itself. Even the rawer DTI measure of white matter integrity, functional anisotropy (FA), has been reported as globally greater in males in some studies (e.g., Cox et al., 2016) but females in others (reviewed in Kochunov et al., 2015). Notably, sex differences in FA are largely eliminated after controlling for ICV (Takao et al., 2014). Thus, it is unsurprising that the structural connectome, which in most cases is derived from DTI measures, fails to exhibit reliable s/g differences, especially when controlling for brain size.

5.2. Sex/gender effects on the functional connectome

Turning to functional measures of the human connectome, we identified 15 large or highly-cited studies (Suppl. Table 3B) reporting male-female comparisons. All of these originate with measures of resting state or "intrinsic" fMRI—that is, blood-oxygen level detection (BOLD) in awake subjects who are lying still in the scanner, without a specific mental task to complete. However, their wide variety of analytic approaches and lack of standardized dependent measures (Rajamanickam, 2020) has resulted in a disparate collection of findings with little consensus on s/g difference in the functional connectome.

Thus, two of the earliest and smallest ($N < 50$) studies reported no (Weissman-Fogel et al., 2010) or very small (Bluhm et al., 2008) male-female differences in resting connectivity based on independent component analysis (ICA) of fMRI data. Where small differences were found by Bluhm et al. (2008) they were confined to the default mode network (DMN), a collection of largely midline structures thought to underlie the brain's "idling" state. DMN connectivity was higher in females in Bluhm et al. (2008), however, the specific circuits varied according to analysis method: ICA resulted in s/g differences in the bilateral superior frontal gyri whereas seed-based connectivity found differences in the medial prefrontal cortex.

Around the same time, Biswal et al. (2010) published a massive analysis of the functional connectome in a study of 1,414 adults collected from 35 different international research centers. Using three distinct statistical methods (seed-based, ICA, and frequency-domain analyses), they found small s/g differences in the connectome, with some brain areas showing stronger connectivity in males and others stronger in females. For any given brain region, Gaussian plots showed more than 80 % overlap in connectivity between male and female groups. Importantly, supplemental data showed that correspondence between the three statistical analyses depends on exactly which of the six seed areas or 20 principal components are analyzed. For example,

although the primary article shows higher connectivity in the posterior cingulate cortex (PCC) of females across all three methods, supplemental figures showed either no sex difference or higher PCC connectivity in males across two of six seed regions and 14 of 20 ICA components.

S/g effects were only a minor focus of Biswal et al. (2010), but its unprecedented size made it a landmark study that is highly cited, even if few consult the supplemental data. Thus, in their review of gender effects on brain connectivity, Gong et al. (2011) cite Biswal as finding generically higher female connectivity in the PCC, medial prefrontal, and inferior parietal cortex, but do not recognize that these findings are tied to specific seed regions or independent components.

Such methodological details are important, and may explain the diversity of s/g findings in ten comparable resting connectome studies carried out since Biswal et al. (2010). Other variation may be attributable to smaller sample sizes, such as contrasts between Tian et al. (2011), who found no s/g differences in their study of young adults, versus Filippi et al. (2013); Scheinost et al. (2015), and Smith et al. (2014), who all report differences, but each in a different set of brain regions (Suppl. Table 3B).

Among the nine largest studies (>493 participants), consensus can be found among six that report stronger DMN connectivity in females (Allen et al., 2011; Biswal et al., 2010; De Lacy et al., 2019; Ritchie et al., 2018; Tomasi and Volkow, 2012a; Zhang et al., 2018). However, there are many discrepancies, even between studies that use the same 1000 Functional Connectome database. For example, Tomasi and Volkow (2012a) found higher connectivity densities across many regions in females, but no regions of higher connectivity in males. This contrasts starkly with Biswal et al. (2010) and Zuo et al. (2012), who used the same database but found some nodes of higher connectivity in females and other nodes in males.

Other disagreement concerns s/g difference in the degree to which connectivity is locally clustered versus integrated across distant brain hubs. Using graph theory methods, which have been applied to network organization in fields as diverse as sociology and transportation, neuroscientists have concluded that the human brain is organized in small world topology, comprised of clustered local nodes with high levels of nearby connections and a short distance between any two nodes. Small-world organization characterizes both male and female brains (Sun et al., 2015; Zhang et al., 2016), but there is disagreement about the degree of connectivity in local versus global networks in males and females. Thus, several large studies find evidence for greater within-network connectivity in female brains and between-network connectivity in male brains, indicating greater overall integration in males and segregation in females (Allen et al., 2011; Satterthwaite et al., 2015; Tomasi and Volkow, 2012a). However, a more recent analysis involving 494 participants reports the opposite result, greater segregation or, local clustering, in male brains (Zhang et al., 2016). Moreover, the finding by Satterthwaite et al. (2015) is directly opposite that of Ingalkhalikar et al. (2014) from the same laboratory, who used the same dataset but a structural analysis (Suppl. Table 3A) to find greater modularity or network segregation in males.

With regard to the developing brain (Suppl. Table 3C), existing research is divided on whether male/female connectome differences either do (Wu et al., 2013; Zielinski et al., 2010) or do not (Geng et al., 2017; Nota et al., 2017; Solé-Padullés et al., 2016) emerge prior to puberty. Notably, the only one of these five to include infants found no s/g difference in either structural or functional connectivity up to age two (Geng et al., 2017).

In summary, the current body of data on s/g difference in the human connectome is conflicted and inconclusive. Although several large studies agree in identifying a pattern of higher DMN connectivity in females (but see Tunç et al., 2016), there is poor agreement about which components of the diffuse DMN network this includes. Female brain networks appear to be more efficient, but it is debated whether this results in more or less integration across networks. Much of the discrepancy in s/g findings is likely due to the wide range of structural

and functional methods used to define and measure the connectome (Reid et al., 2017; Rajamanickam, 2020). Another issue is the degree to which spatial variability muddies the automated pipelines used to construct functional connectomes (Bijsterbosch et al., 2018)—that is, individual differences in structure may distort findings about functional connectivity, especially considering differences in hemispheric shape between males and females (Zilles et al., 2001). Finally, brain size remains a key covariate that has been inconsistently applied in analyses of both the structural and functional connectome. Generally speaking, studies of the structural connectome that correct for individual brain size find smaller or absent s/g effects (Dennis et al., 2013; Duarte-Carvajalino et al., 2012; Hänggi et al., 2014; Lv et al., 2010; Mechelli et al., 2005) compared to studies that do not account for brain size (e.g., Ingallalikar et al., 2014). Total brain volume also appears to contribute to s/g differences in network efficiency (Yan et al., 2011). It is less clear how brain size may affect measures of the functional connectome; although network architecture is plausibly sensitive to absolute volume, three studies that addressed the issue reported no effect of covarying for total gray matter or ICV (Filippi et al., 2013; Wang et al., 2012; Zhang et al., 2016) on s/g difference in functional connectivity (but see Zhang et al., 2018).

6. Sex/gender prediction using multivariate statistical learning algorithms

Despite the lack of consensus about specific connectome differences, another approach that deploys the same massive datasets has been hailed as an important advance toward understanding male/female brain difference. These are studies using a variety of machine learning

tools to predict individuals's/g based on brain MRI (or in one case, EEG) data. Here again, the precise methods vary across studies, but all involve plugging complete imaging datasets, involving hundreds or more variables per brain, into an artificial intelligence algorithm and asking it to find features that reliably predict whether a given brain belongs to a male or female participant.

We identified 12 independent studies of this type (Table 7), nearly all reporting s/g prediction accuracy in the 80–90 % range. While impressive, most of this discrimination appears to be based on brain size, which as we've seen, correlates strongly with regional volumes, surface area, DTI, and other connectivity measures. Although one of the earliest studies (Wang et al., 2012) claimed to retain 91 % accuracy in a subsample of 35 male and 35 female brains matched for TBV, this and another small study (Feis et al., 2013) failed to divide their sample into training and testing cohorts, a minimal requirement for measuring classification accuracy.

The remaining 10 studies all used suitably large samples ($N > 673$) and divided them into training and testing groups, but only three of them compared prediction accuracy after controlling for total brain size. The importance of this control was first shown by Chekroud et al. (2016) who obtained 93 % s/g prediction accuracy based on a multivariate morphometric dataset, but this declined to 70 % when head size was regressed out. (Note that with only two categories to sort, chance accuracy is 50 %.) More recently, Sanchis-Segura et al. (2020) dissected the contribution of brain size to s/g prediction accuracy. Using 12 different algorithms and matched numbers of males and females, they showed that an average 82 % accuracy declined to 60 % when regional volumes were properly corrected using either residual or power-proportions methods (See Section 2.3). Sanchis-Segura et al.

Table 7
Multivariate sex/gender prediction.

First author (year)	#M/F (source)	Mean age	Multimodal variables	Prediction accuracy	Brain size control	Key discriminating features (bold) & other comments
Wang et al. (2012)	70/70 (1000 FCP)	20.9	Combination of GM density & rsfMRI regional homogeneity	89 % [†]	Yes: matched TBVs, n = 70; 91 % accuracy	Discriminating GM densities scattered across cortex & cerebellum. For rsfMRI, relatively greater synchrony in males' R vs. females' L hemispheres
Feis et al. (2013)	54/67	25.0	Morphometry: T1, T2 and FA signals	96 % [†]	No	Mostly rostral to post-central sulcus & favoring F volumes (bilateral frontal and L. perisylvian)
Satterthwaite et al. (2015)	362/312 (PNC)	15.7	Multivariate correlation of rsfMRI and cognitive profile	68–71%	No	Greater between-module connections in M and within-module connections in F. Functional connectivity predicted sex/gender better than cognitive measures.
Chekroud et al. (2016)	663/903	Range: 19–35	T1, cortical thickness & subcortical vols.	93 %	Yes, declined to 70 %	Not analyzed
Joel et al. (2018)	360/495 (Connectomes+)	27.7	VBM, supervised clustering	77 %	No	Classification accuracy declined to near chance in 2 out of 3 external samples.
Sepehrband et al. (2018)	469/498 (PNC)	14.7	Cortical thickness, area, vol., curvature	77 % [†]	No	Thicker left medial occipital & bilateral angular gyri in F
van Putten et al. (2018)	746/562	43.4	26-channel EEG	81 %	No	No distinct spatiotemporal patterns observed. Fast beta activity (20–25 Hz) main distinguisher
Zhang et al. (2018)	366/454 (HCP)	28.8*	Resting fMRI connectivity	87 %	Yes, declined to 70 %	"Significant FC features were scattered across the brain"
Anderson et al. (2019)	1014/286 (incarcerated)	31.3	Regional gray matter densities & volumes	93 %	No	Orbitofrontal and frontal pole larger in F; anterior medial temporal regions larger in M
Luo et al. (2019)	507/606 (HCP, S1200)	Range: 22–37	Cortical thickness & surface area plus rsfMRI connectivity	97 %	No	Frontal lobe, especially superior frontal gyrus, plus DMN (rsfMRI) most important for classification accuracy
Xin et al. (2019)	490/575 (HCP)	Range: 22–36	3D convolutional network based on DTI and 298 ROIs	93 %	No	GM: L. precuneus, post-central, subgenual cingulate & R. orbital. WM: middle cerebellar ped., corpus callosum genu, R. sup. & ant. corona radiata, L. ant. limb internal capsule
Sanchis-Segura et al. (2020)	444/444 (HCP) 171/171 (UJI)	28.6 22.0	Cortical & subcortical GM volumes	82 %	Yes, declined to 60 %	Accuracy lower in external validation (UJI) sample, but both HCP test and UJI were ~60 % with ICV residual correction

Abbreviations as above plus: 1000FCP (1000 Functional Connectomes Project); PNC (Philadelphia Neurodevelopmental Cohort); HCP (Human Connectome Project); UJI (see Sanchis-Segura et al., 2019); rsfMRI (resting state functional MRI); VBM (voxel-based morphometry); DMN (default mode network).

*Significant sex difference ($M = 28.0 \pm 3.7$, $F = 29.4 \pm 3.6$ years).

†Accuracy calculated using a single sample (no independent training/testing samples).

‡Prediction accuracy varied by brain atlas.

(2020) further tested their algorithms against an external population and found that accuracy fell even below 60 %.

The third study to control for brain size was by Zhang et al. (2018), who again found s/g prediction accuracy declined substantially, from 87 % to 70 %, when they regressed out the contribution of total brain and CSF volume. Importantly, this study was not based on structural measures, but achieved its classification using functional connectivity derived from rsfMRI. This finding demonstrates the importance of correcting for brain size even when assessing individual differences in functional measures, which as we saw in the previous section is not current practice. But it is not surprising that brain size influences such connectivity given the underlying features such as callosal area and inter-/intra-hemispheric fiber ratio that correlate with ICV and TBV.

Beyond the confound of brain size, the 12 studies in Table 7 differ strikingly in features found to be most important for s/g classification accuracy. Of course, one would not expect similar features to emerge between studies using qualitatively different data, such as rsfMRI activity versus regional gray matter volumes. But even among studies that relied exclusively on structural measures, we see a lack of replication among the brain regions identified as most important for male/female classification across studies. Thus, Sepehrband et al. (2018) found the strongest discriminators to lie within the left medial occipital and bilateral angular gyri. By contrast, Anderson et al. (2019) reported that orbitofrontal and anterior medial temporal regions weighed more heavily in their s/g prediction. Luo et al. (2019), which reported the highest prediction accuracy, found the superior frontal gyrus to contribute most to classification accuracy, whereas Xin et al. (2019), using the same dataset, list a smattering of cortical GM and subcortical WM structures as contributing more strongly. Notably, Zhang et al. (2018) were unable to identify specific components of functional connectivity (FC) that differed most importantly between males and females, concluding “It is hard to distinguish the brain regions with more important FC features for gender prediction based on the distribution of feature weight plot... as the significant FC features were scattered across the brain.”

Collectively, these machine learning studies demonstrate that it is possible to predict with good confidence from MRI data whether an anonymous brain is from a male or female individual. However, this discrimination is largely based on brain size and there is no agreement about local features that are most important for distinguishing male versus female types. The lack of hallmark “male” versus “female” brain features is likely because each algorithm was custom-developed for its particular dataset. Thus far, the only two s/g prediction studies to test their algorithms on external populations both found their accuracy to drop to near chance levels (Joel et al., 2018; Sanchis-Segura et al., 2020), indicating the classifiers are not tapping into universal features that discriminate male from female brains. These findings challenge the notion that there exists a discrete set of variables that capture core differences between male and female brains across the human species. Indeed, a recent comparison of s/g differences in Chinese vs. American samples found almost no overlap in brain regions that discriminate males from females based on sulcal depth, cortical thickness and myelin content (Yang et al., 2020a). In other words, s/g discriminatory features are highly population-specific, probably reflecting a combination of local genetic and environmental influences. Moreover, this discrimination is strongly influenced by brain size (Sanchis-Segura et al., 2020) and probably also cranial shape (Zilles et al., 2001), which has been inadequately studied.

7. Task-based fMRI

Ultimately, the goal of research on s/g brain difference is not to predict whether individuals are male or female (there are much easier ways to do this), but to understand the neural basis of behavioral differences between males and females. In particular, the most common pretext stated in studies of brain s/g difference is to better understand

and treat disorders such as autism, dementia, depression, anxiety, ADHD and dyslexia (Fig. 1) that differ in prevalence between males and females (e.g., Kaczurkin et al., 2019; Pallayova et al., 2019).

Thus, a more fruitful approach for detecting meaningful s/g brain differences is to focus on specific mental tasks, especially those that differ most reliably between men and women at the group behavioral level. In fact, task-based fMRI studies of s/g difference pre-date connectome studies by more than two decades. Given that s/g differences in behavior and personality have been thoroughly documented (Hyde, 2014; Zell et al., 2015), the chance of detecting reliable male/female brain differences would be seemingly much greater by focusing on specific tasks. A wide array of behaviors has been assessed for fMRI s/g differences, ranging from sexual arousal (Mitricheva et al., 2019) to mathematical processing (Keller and Menon, 2009) and humor appreciation (Azim et al., 2005). Here we focus on the three domains that have been most extensively studied: language (Tables 8A, 8B, 8C, 8D), visuo-spatial (Tables 9A, 9B, 9C), and emotion (Table 10) processing.

7.1. Language processing

One of the first reports of an fMRI activation difference of any kind between males and females was by Shaywitz et al. (1995; Table 8D), as mentioned in Section 4. This study found that males exhibited more left-lateralized activity of the inferior frontal gyrus (i.e., Broca's area) than females during performance of one out of three verbal tasks. However, this was not supported by subsequent studies (e.g., Frost et al., 1999) or two meta-analyses (Sommer et al., 2004, 2008). At the behavioral level, the claim that males are more lateralized for language was initially based on clinical and dichotic listening studies (McGlone, 1980), but the clinical argument has been disproven through better-controlled studies of aphasia patients: s/g was not a reliable predictor of aphasia incidence, severity or recovery in two analyses of large patient populations by Pedersen et al. (1995, 2004). Similarly, based on a systematic review of post-stroke aphasia, Plowman et al. (2012) concluded that the “gender does not appear to significantly impact incidence of aphasia, aphasia type or recovery patterns.” The fact that language ability is comparably affected in males and females following stroke argues against the claim that the neural circuitry for language is organized in a meaningfully different way between sexes.

This conclusion is supported by the consensus of 33 fMRI language-processing studies cataloged in Tables 8A, 8B, 8C, 8D, where the range of findings is immense. Of these, 16 studies reported brain areas with higher activity in males; another 16, partially overlapping set of studies found higher activity foci in females. And the foci implicated in each group range across the cortical mantle, subcortical structures, and both hemispheres, with almost no replication across studies.

An obvious confound is the variety of language tasks employed. However, even when we narrow down the studies to those employing the same task—verbal fluency (Table 8A), verb generation (Table 8B), and verbal working memory (Table 8C)—there is lack of any consensus.

Verbal fluency has the distinction of being the one language task to show the largest s/g difference at the behavioral level (Hyde and Linn, 1988). Verbal fluency is typically assessed by asking subjects to rapidly name as many words as possible that begin with the same letter, such as “B,” or as many words as possible that fit into a specific object class, such as “4-legged animals.” According to the meta-analysis by Hyde and Linn (1988), women out-perform men in verbal fluency with a d-value of 0.33. Of the six fMRI studies (including 7 independent samples) we identified that explicitly addressed s/g difference during verbal fluency tasks (Table 8A), three reported no difference in brain activation (Gaillard et al., 2003; Halari et al., 2006; Weiss et al., 2003a). Three others found some evidence of male/female divergence, however, the brain locations and direction of difference were entirely different among them (Beacher et al., 2012; Gauthier et al., 2009; Schlosser et al., 1998). So even though verbal fluency is the most reliable of the linguistic differences between males and females, there is no consensus about a

neural basis for this difference and more evidence for *similar* activation in men and women in at least one structure, the left inferior frontal gyrus.

Another task frequently used to look for s/g difference is verb generation—asking subjects to silently think of a single verb that goes with a particular noun, such as “eat” when they see an image or the written word “cookie.” Here again, there is virtually no consensus among the six fMRI studies that analyzed their activation data for male/female difference (Table 8B), with three finding areas of greater activation in males, two finding areas of greater activation in females, and no overlap between studies in specific foci.

A last language task that is frequently used to assess s/g difference in neural processing is verbal working memory. Five language-processing studies performed fMRI during some type of verbal working memory task (Table 8C). Among these, four found areas of greater activation in females (Goldstein et al., 2005; Haut and Barch, 2006; Speck et al., 2000; Valera et al., 2010), two found areas of greater activation in males

(Speck et al., 2000; Valera et al., 2010), and one found no s/g differences in fMRI activity (Schmidt et al., 2009). Once again, none of the sex-specific regions were replicated between studies.

The lack of reproducibility in these findings strongly implicates false discovery. As discussed above, David et al. (2018) concluded from their meta-analysis that the fMRI literature on brain s/g difference demonstrates “excess significance”—that is, more statistically significant results than would be predicted from the summary effect size. This indicates the presence of publication bias and likely, false positive findings in some subset of published studies. This is especially likely considering the small size of most of these studies, well below the number of participants needed to adequately power the detection of small-to-moderate sized differences (Button et al., 2013; Carp, 2012). But even when we look at the two largest studies in this collection—both analyses of verb generation in similarly-aged children from the same laboratory—one reported greater female activation in the precuneus and middle frontal gyrus (Schmithorst and Holland, 2006), whereas the

Table 8A
fMRI sex/gender differences during verbal fluency.

First author (year)	N	Mean age (range)	Performance difference?	Areas of M>F activation	Areas of F>M activation	Areas of similar activation in M & F
Schlösser (1998)	12	23	Yes	None	Medial frontal and R. orbitofrontal	L. prefrontal & dorsal parietal; R. cerebellum; Deact. bilateral dorsolateral & medial parietal
Gaillard (2003)	16	10.2	n.r.	None	None	L. IFG, MFG, medial frontal, parietal
	29	29.2	n.r.	None	None	L. IFG, MFG, medial frontal, parietal, temporal; R. MFG; Bilateral occipital
Weiss (2003a)	20	25.3	No	None	None	BA 46, 47, 32 (cingulate) and cerebellum
Halaris (2006)	19	25.3	Yes	None	None	IFG, caudate
Gauthier (2009)*	44	19.9	Yes	L. ITG, ACC, PCC and cerebellum; R. SFG & dlPFC, ACC, lingual	None	L. SPL, IPL, thalamus; R. SMA; Bilateral IFG, inf. visual, cerebellum & basal ganglia
Beacher (2012)*	61	31.5	No	L. caudate tail; R. parahippocampal	None	L. IFG and cingulate

Table 8B
fMRI sex/gender differences during verb generation.

First author (year)	N	Mean age (range)	Performance difference?	Areas of M>F activation	Areas of F>M activation	Areas of similar activation in M & F
Drobyshevsky (2006)*	31	40.9	None	None	None	L. IFG, STG, MTG and parietal cortex; Bilateral cingulate
Plante (2006)*†	205	(5-18)	Prosody only	None	None	Frontal and temporal ROIs
Gizewski (2006)*	26	29	n.r.	L. medial temporal & precentral cortex	None	L. inferior frontal, medial temporal, fusiform cortex, thalamus, and lentiform nucleus
Schmithorst (2006)*	323	11.8	No	None	precuneus and MFG	Broca's area, MFG, MTG, anterior cingulate & precuneus
Rumberg (2010)*	36	(18-43)	No	L. STG (compared to women in luteal phase only, not menstrual)	L. IFG (only luteal phase women)	L. IFG, STG, MTG; bilateral cuneus
Allendorfer (2012)*‡	40	40.3	No	R. MFG, SFG, caudate, anterior cingulate (only noun-verb task)	None	L. MFG and MTG; Bilateral IFG, SFG, cingulate, cuneus, insula, caudate

Table 8C
fMRI sex/gender differences during verbal working memory.

First author (year)	N	Mean age (range)	Performance difference?	Areas of M>F activation	Areas of F>M activation	Areas of similar activation in M & F
Speck (2000)	17	33.7	n.r.	R. parietal & lateral PFC	L. parietal & lateral PFC	lateral PFC, parietal cortex, SMA, caudate
Goldstein (2005)*	14	33.1	No	None	L. dorsolateral PFC; R. Broca's & orbitofrontal	L. Broca's; R. posterior parietal cortex
Haut & Barch (2006)*	110	36.5	No	None	R. occipital	L. inferior frontal, superior parietal and middle temporal gyri
Schmidt (2009)*	50	33.7	No	None	None	Bilateral SFG, MFG, IFG, IPL
Valera (2010)*	63	34.5	No	R. inferior parietal lobule	L. insula, putamen & pallidum	n.r.

Table 8D

fMRI sex/gender differences during other verbal tasks.

First author (year)	N	Mean age (range)	Task	Perform. diff.?	Areas of M>F activation	Areas of F>M activation	Areas of similar activation in M & F
Shaywitz (1995)	38	26.3	Letter recognition, rhyming, semantic category	No	None	R. IFG	Extrastriate (spelling task); IFG, STG, MTG for rhyming & semantic
Frost (1999)	100	23	Semantic Monitoring/Association Task	No	None	None	L>R prefrontal, temporal, angular, retrosplenial, thalamus; Bilat. cerebellum
Kansaku (2000)	47	(20-47)	Story Processing	Yes	L. MTG	R. TPC	L. TPC; R. MTG; Bilateral IFG & STG
Kocak (2005)*	12	(28-49)	Passive text listening + noise	N/A	STG, MTG and IFG, to high vs. low scanner noise	None	L. STG, superior MTG, STS, IFG
Plante (2006)*†	205	(5-18)	Vocab.-Picture ID, Story Proc., Syntactic Prosody	Prosody only	None	L. IFG & bilateral STG for Story Processing; L. IFG prosody only	Frontal and temporal ROIs
Clements (2006)*	30	(19-35)	Phonological Test: Rhyming analysis	No	None	None	L. IFG
Kaiser (2007)*	44	(20-36)	Silent Free Narration	n.r.	R. Broca's (less lateralized)	L. Broca's (more lateralized)	Broca's, Wernicke's, visual cortex, premotor, prefrontal, cingulate
Burman (2008)*	62	15-Sep	Orthographic (spell) & phonologic (rhyme) judgment	Yes	None	Bilateral IFG & STG, L. fusiform & caudate; R. MeFG & cingulate	L. MFG, MTG, MeFG, cingulate, occipital, parahippocampal, IPL; Bilateral insula & precentral
Harrington (2008)*	40	35.7	Multiple tasks	n.r.	L. inferior frontal orbitalis	R. insular cortex	Bilateral IFG, TPC, MTG and insular cortex
Konrad (2008)*	24	32	Synonym generation	No	L. precentral & MFG	None	Bilateral IFG, MFG, SMA, cingulate; L. lingual
Weis (2008)*	28	27.1	Lexical decision-making (word-matching)	Yes (RT)	L. MFG, medial frontal, cingulate; R. IFG, angular, superior parietal	L. SFG	L. IFG
Garn (2009)*	26	23.6	Object naming	n.r.	None	None	Bilateral fusiform; L. medial SMA, superior parietal, thalamus, middle cingulate
Ihnen (2009)*	26	24.9	Word Generation	No	17 regions with significant Sex x Time interaction	None	L. IFG & cingulate; R. MFG, MTG, parahippocampal, claustrum; Bilateral STG, pre/post-central
Ihnen (2009)*	20	25.3	Word/Nonword Reading	No	Only 1 of 17 regions found in word generation task	13 regions with significant sex*time interaction	L. post-central; R. STG; Bilateral MTG, cingulate, posterior insula, inf. occipital
Larsen (2010)*	10	(22-30)	Word memory test	No	None	None	R. dorsal anterior cingulate, SMA; Bilateral lingual, MFG, anterior insula, SPL
Perrone-Bertolotti (2011)*	22	26.5	Simple vs. complex Grapheme-to-phoneme	Sex x Task	None	L. lingual gyrus (simple task only)	L. IFG, STG, ITG and premotor
Kana (2012)*	36	n.r.	Literal & idiom sentence comprehension	No	No foci, but greater connectivity between L. IFG and MTG	L. MFG and SFG; R. postcentral; bilateral precuneus	L. IFG, SFG; R. MFG, SPL; Bilateral occipital
Tschernegg (2017)	40	26.5	Letter matching (single task)	No	R. postcentral, SPL, PT, precuneus; L. PFC, sup. occipital, STG, orbito-frontal, amygdala, hippocampus	None	n.r.

Sex/gender difference in fMRI activation during A) verbal fluency, B) verb generation, C) verbal working memory and D) other tasks. "Performance difference" (Perform. Diff.) indicates whether a significant sex/gender behavioral difference was observed to the task performed during scanning, measured as either accuracy or reaction time. Abbreviations as above plus: ACC (anterior cingulate g.); IPL (inferior parietal lobule); MeFG = medial frontal gyrus; PFC (prefrontal cortex); PT (planum temporale); RT (reaction time); SMA (supplemental motor area); SPL (superior parietal lobule); STG (superior temporal g.); TPC (temporoparietal cortex).

*Study was included in meta-analysis by David et al. (2018).

†Same study, multiple tasks (Tables B and D).

‡Activity measured during both overt verb generation and noun-verb association tasks.

Citations not included in article text: (Allendorfer et al., 2012; Burman et al., 2008; Dobroshevsky et al., 2006; Garn et al., 2009; Gizewski et al., 2006; Harrington and Farias, 2008; Kaiser et al., 2007; Kana et al., 2012; Kansaku et al., 2000; Kocak et al., 2005; Konrad et al., 2008; Larsen et al., 2010; Perrone-Bertolotti et al., 2011; Rumberg et al., 2010; Tschernegg et al., 2017; Weis et al., 2008).

other found no s/g difference in fMRI activation (Plante et al., 2006; Table 8D).

The possibility that fMRI sex difference research is vulnerable to false discovery was tested explicitly by Ihnen et al. (Table 8D). Although this study identified 17 specific brain regions that differed in activity between males and females, only one of the 17 foci generalized between two closely-related language tasks. Moreover, they found a comparable number of group-level activation differences when the same subjects were divided into random groups, not separated by s/g. In other words, Ihnen et al. (2009) were able to generate group-level differences by creating fake groups, a control that few other studies have employed. Based on these findings, along with the "grab bag" of results they review from prior investigators, Ihnen et al. (2009) concluded: "our results do not support a particular model of sex differences in language processing, and they overwhelmingly do not specifically corroborate previous fMRI studies of sex differences in language."

In summary, and in spite of the quarter-century of research in this area, there is no reliable evidence that male and female brains process language differently. A similar conclusion was reached by other reviewers (Kaiser et al., 2009; Wallentin, 2009) who surmised that this literature is highly subject to publication bias; studies or post-hoc analyses that find significant s/g effects are likelier to find their way into

print than analyses that failed to find a male/female difference. And even among the published positive findings, the foci identified as differentially active between males and females rarely replicate between studies (Tables 8A, 8B, 8C, 8D). Given that many of these are now appreciated to be statistically under-powered, it is safe to conclude that fMRI research has failed to uncover meaningful differences in the circuits males and females use to perceive, process, or generate language. A similar conclusion was reached by Etchell et al. (2018) in their systematic review of language-related brain activation in girls vs. boys.

7.1.1. Spatial processing

In contrast with verbal performance, s/g differences in spatial tasks tend to favor males and the effect sizes are larger, at least between adult men and women (Maeda and Yoon, 2013). Spatial cognition has been partitioned into three domains: visualization, perception, and mental rotation. Early meta-analyses by Linn and Peterson (1985) and Voyer et al. (1995) found that males out-perform females in all of these domains, although the difference in spatial visualization is very small ($d = 0.13$ to 0.19) and not significant before 18 years of age. S/g differences in the other domains are moderate-sized: $d = 0.44$ for spatial perception and $d = 0.56$ for mental rotation (Voyer et al., 1995), the latter in close agreement ($d = 0.57$) with a more recent meta-analysis by

Maeda and Yoon (2013). Recently, Nazareth et al. (2019) conducted a meta-analysis of male/female difference in spatial navigation, an important real-world application of this form of cognition and found a pooled effect size averaging $d = 0.36$, favoring males. For all domains, the magnitude of the s/g difference in spatial cognition was smaller in children (Lauer et al., 2019; Nazareth et al., 2019; Voyer et al., 1995), and the preponderance of evidence indicates it is not present in infancy and early childhood (Frick et al., 2014).

In spite of reliable group differences between men and women in spatial skill performance, neither structural nor functional brain assessments have demonstrated a reproducible neural basis for it. The earliest studies used EEG and evaluated both children and young adults.

In an initial study, Roberts and Bell (2000) found greater activation in left parietal and posterior temporal areas for men, compared to women, during a 2-dimensional mental rotation task (MRT), but no neural difference between 8-year-old boys and girls. However, the same authors found no s/g difference in EEG activity at either age using a more challenging, 3-dimensional MRT (Roberts and Bell, 2000, 2003), a paradoxical finding considering that this 3D task incurred the usual male performance advantage.

As MRI studies began to multiply, it seemed likely that the neural substrate for s/g differences in spatial cognition would be better elucidated. We did not conduct a fully systematic review of such research, but did collect 39 such studies (Tables 9A, 9B, 9C) through a combination of

Table 9A

Structural MRI correlates of sex/gender spatial skill difference.

First author (year)	#M/#F	Mean age	Brain measure	Task	Perform. Diff.?	Correlate of M performance	Correlate of F performance	Similar correlates in M and F
Koscik (2009)	38/38	26.7	Cortical vols. & SA, corrected for age and total vol. or SA	MRT	Yes	Parietal lobe SA	Negative correlation with L. parietal GM:WM ratio	No sex difference or correlation between parietal GM or WM volumes and MRT performance.
Hänggi (2010)	18/25	22.3	GM and WM volumes, controlled for SNP of BDNF	Block design	No	Parietal WM (R>L)	Parietal GM (L>R)	GM: L. IPL, precuneus & angular; Bilat. SPL, pre-central, post-central WM: Bilat. angular; L. IPL; R. SPL, precuneus, postcentral, STG, SMA
Wei (2016)	192/239	19.9	GM volumes, controlled for TBV	MRT	Yes	Greater GM volume in R. anterior hippocampus correlated with MRT performance, including male advantage.	GMV of calcarine and anterior hippocampus	
Martinez (2017)	44/58	19.8	DTI (structural connectome)	Battery of 7 spatial tasks, including 2 MRT	No for overall spatial battery	None	None	Greater inter- to intra-hemispheric connectivity ratio correlated with better performance in both sexes. This ratio is higher in women, but is a function of brain size, not sex per se

Abbreviations as above, plus: BDNF (brain-derived neurotrophic factor); MRT (mental rotation task); SNP (single nucleotide polymorphism); SA (surface area).

Table 9B

Mental rotation task fMRI sex/gender differences.

First author (Year)	#M/#F	Mean age	Perform. Diff.?	Areas of M>F activity	Areas of F>M activity	Similar activation M & F
Jordan (2002)	10/14	23.2	No	Bilat precentral; R. POS; L. IPS	Bilat. IPS, SPL, IPL, ITG; L. motor	L. SPL; premotor, SMA
Weiss (2003b)	10/10	Univ.	Yes	Bilateral IPL	Bilat. SPL, R. IFG	IFG, MFG, SFG, cingulate, IPL
Seurinck (2004)	11/11	25.4	No accuracy; Yes RT	Lingual	Ventral premotor	SPL, dorsolateral PFC, extrastriate, occipital
Kucian (2005)	10/10	25.9	No	None	Bilat. precentral; R. MTG, IFG	Bilat. fusiform; L. IPL, precentral; R. IFG, IPS, middle occipital
Butler (2006)	12/13	29.3	No	Ventral anterior cingulate (vACC)	vACC suppression	n.r.
Halari (2006)	9/10	25.3	Yes	L. MTG; R. angular	none	R. SPL; bilat. middle occipital
Hugdahl (2006)	6/5	30	No	R. SPL	R. IFG	Bilateral SPL, bilateral IFG
Gizewski (2006)	12/14	29	n.r.	R. medial prefrontal, precentral; Bilateral IPL	R. ITG, medial temporal, STG, fusiform	Bilateral IPL, SPL, medial & inferior prefrontal, pre- and postcentral
Kucian (2007)	10/10*	10.5	No	None	None	Bilat. SPL; R. fusiform
Schöning (2007)	12/12	32	Yes	L. IPL, fusiform and lingual g. (vs. women in either menstrual phase)	Luteal phase: L. STG, SFG, R. caudate. Follicular: L. precentral; R. MFG & post-central; cerebellum	Bilat. SPL, IPL, medial & lateral frontal cortices, occipital lobe, and cerebellum
Christova (2008)	4/4	26.8	No	None, but more R-lateralized	None, but less lateralized	Widespread parietal, frontal, occipital foci
Carrillo (2010)	23/19†	32	Yes	L. precuneus & lingual	L. postcentral	Bilat. parietal, occipital and lateral frontal
Hattemer (2011)	10/10	26.1	Yes	R. precuneus, MFG, SFG, POS, lingual; L. precentral	R. precentral	Bilat. SPL, POS, MFG, SFG, insula, thalamus, midbrain, cerebellum; R. IFG
Beacher (2012)	16/16	30.4	No	Left cerebellum	None	Bilat. SPL, frontal & infero-temporal cortex
Hoppe (2012)	8/9	16.5	Yes	L. IPL, precentral, precuneus; R. postcentral	L. MFG	Bilat. IPL
Semrud-Clikeman (2012)	20/20	N/A	No	Right IFG & LFG; L. precuneus, cuneus, PCC, middle occipital	None	Bilat. MFG & IPS; L. precuneus
Burke (2016)	20/21†	16.1	No	None	R. frontal and L. parietal	Bilat. parieto-occipital and frontal networks
Van Hemmen (2016)	30/29‡	31.6	No	Bilateral IPL	None	Bilat. parietal, occipital, & frontal (precentral, SFG, MFG) lobes
Tschernegg (2017)	20/20	26.5	No	None	L. IFG, pars triangularis (single task)	n.r.

Abbreviations as above plus: IPS (intraparietal sulcus); POS (parieto-occipital sulcus).

*Child sample only. Adult sample reported in Kucian et al. (2005).

†Cis-gender participants.

‡Healthy controls.

review article citations, Google Scholar top citations, and mining the meta-analysis by [David et al. \(2018\)](#). The vast majority of studies we collected assessed fMRI activation. In addition, four studies reported structural MRI correlates of spatial cognition, and one compared the connectivity patterns of spatial processing between men and women based on rsfMRI activity ([Boccia et al., 2017](#)).

7.1.2. Structural brain correlates of male-female spatial skill differences

Taking the four structural studies first ([Table 9A](#)), the earliest ([Koscik et al., 2009](#)) focused on the parietal lobe and reported that better MRT performance was correlated with parietal lobe surface area in men; in females, however, it was the proportion of parietal GM/WM that correlated with MRT performance, in this case predicting *poorer* performance. [Hänggi et al. \(2010\)](#) conducted a similar study using a block design task, but obtained nearly opposite results: performance was correlated with parietal WM volume in males and parietal GM volume in females. The third and largest of the structural studies was by [Wei et al. \(2016\)](#) who reported no differences in the parietal lobe but a significant correlation between right anterior hippocampal GM and MRT score, with males exhibiting a larger anterior hippocampus than females (after controlling for brain size). These authors concluded that right anterior hippocampal GM is the “neurobiological substrate for the sex difference in 3D mental rotation.”

The fourth and most recent study to assess a structural correlate of spatial skill performance found no s/g differences. [Martinez et al. \(2017\)](#) compared DTI connectivity measures in men and women to their performance on a battery of spatial skills and observed that spatial skill

performance correlated with the ratio of inter- to intra-hemispheric connectivity across all individuals, male or female. Recall that [Ingalhalikar et al. \(2014\)](#) found that this ratio is higher in females, or opposite what would be predicted based on the performance difference. However, [Martinez et al. \(2017\)](#) also found that controlling for either ICV or TBV eliminated the s/g difference in this ratio (similar to [Hänggi et al., 2014](#)). Thus, a greater degree of inter-hemispheric communication appears to help both males and females with spatial skills and, according to [Martinez et al. \(2017\)](#), verbal skills as well.

In summary, extant structural MRI studies of spatial processing circuitry have produced highly conflicting results regarding the neural correlates of s/g difference, and no replications.

7.1.3. fMRI correlates of male-female spatial skill differences

Much more attention has been paid to fMRI correlates of spatial skill s/g difference, where the number of studies is large enough that they can be separated by the type of task performed in the scanner. Of the 34 such studies we identified, the majority (19) used a MRT to search for sex difference in brain activation ([Table 9B](#)). According to two meta-analyses, neither of which assessed s/g effects, mental rotation activates two primary brain areas bilaterally: one centered on the intraparietal sulcus and the other on the medial superior premotor cortex ([Hawes et al., 2019; Zacks, 2008](#)). Among the 19 studies in [Table 9B](#), nearly all report some type of parietal activation, mostly bilateral, in both males and females. Many also report some frontal lobe activation. But the studies range enormously in their findings of s/g difference.

Most of these MRT studies were quite small and only six replicated

Table 9C

Other spatial tasks, fMRI sex/gender differences.

First author (Year)	#M/#F	Mean age	Perform. Diff.?	Areas of M>F activity	Areas of F>M activity	Similar activation M & F
VIRTUAL NAVIGATION TASK						
Grön (2000)	12/12	26.1	Yes	L. parahippocampal; R. precuneus	L. SPL and SFG; R. MFG and fusiform	Bilat. medial occipital, lingual, PCC; L. precuneus; R. SPL, HC, parahippocampal
Ohnishi (2006)	28/28	27.4	No	None	None	Bilat. middle occipital, lingual, fusiform. parahippocampal, HC, PCC, precuneus, cuneus, SPL, IPL; R. MFG
Sneider (2011)	8/7	46.6	No	L. opercularis, amygdala; R. triangularis, orbitalis, precentral	Left STG & ITG	Bilat. HC, parahippocampal, fusiform, precuneus
Persson (2013)	12/12	25.1	Yes	R. HC	None	Bilat. posterior HC
Boccia (2017)	52/56	25.9	n.r.	None	L. retrosplenial cortex-to-L. thalamus connectivity	Bilateral HC, parahippocampal place area, retrosplenial cortex
LINE ORIENTATION TASK						
Gur (2000)	14/13	NR	Yes	R. inferior parietal	L. inferior parietal	Planum temporale
Clements (2006)	15/15	24.2(M) 28.5(F)	No	L. cuneus & PCC	None	Bilateral parietal, precuneus, occipital, fusiform, cerebellum; L. cuneus; R. IFG, MFG, IFT, MTG
Clements-Stephens (2009)	16/16	10.3	No	R. lingual and cerebellum	None	Bilat. SPL, occipital, cerebellum; L. fusiform; R. precuneus, cuneus
TOWER OF LONDON						
Unterrainer (2005)	10/10	22.5	Yes	None	R. hippocampus	R. inferior parietal & dorsolateral PFC
Boghi (2006)	9/9	35.9	No	L. precuneus and SPL	L. precentral; R. SPL & IPL, MFG & postcentral	Bilat. fronto-parietal networks
Staphorsius (2015)	21/24	14.6	No	Bilat. precuneus	None	Bilat. dlPFC & insula; L. SPL, SMA, precentral; R. IPL & opercularis
SPATIAL WORKING MEMORY						
Caldwell (2005)	23/16	16.6	No	L. precuneus, angular & subcallosal	L. MFG & thalamus; R. SFG	n.r.
Schweinsburg (2005)	24/25	14.8	No	R. ACC and SFG	None	n.r.
Squeglia (2011)	31/24	17.9	No	None	None	Bilat. ACC; L.STG, MFG & cerebellum; R. SFG, IFG, MTG
OTHER SPATIAL TASKS						
Kaiser (2008)	12/12	28.2	No	R. precuneus, right IFG	None	R. middle occipital gyrus and SMA
Rubia (2010)	38/25	21	Minor	L. IPL, precuneus & medial temporal	R. ITG, STG, putamen & insula	Bilat. SFG, medial temporal and occipital; R. precuneus, medial IFG, putamen, caudate, PCC

Abbreviations as in prior tables.

Citations not included in article text: ([Boghi et al., 2006; Caldwell et al., 2005; Christova et al., 2008; Gizewski et al., 2006; Jordan et al., 2002; Kaiser et al., 2008; Rubia et al., 2010; Schweinsburg et al., 2005; Squeglia et al., 2011; Staphorsius et al., 2015; Tschernerg et al., 2017; Unterrainer et al., 2005; van Hemmen et al., 2016](#)).

males' performance advantage when subjects were in the scanner. These limitations may explain the wide diversity of results, varying from no s/g difference (Kucian et al., 2007) to greater male activation in the left cerebellum (Beacher et al., 2012), left middle temporal gyrus (Halari et al., 2006), frontal gyri (Hattemer et al., 2011; Semrud-Clikeman et al., 2012), ventral anterior cingulate gyrus (Butler et al., 2006), and lingual gyrus (Carrillo et al., 2010; Seurinck et al., 2004). Conversely, 12 of the studies reported areas of activation that were significantly stronger in females, spread across various frontal, temporal, and parietal sites (Table 9B). Nonetheless, some agreement is suggested in four studies that found greater parietal activation in males versus greater frontal activation in females (Hoppe et al., 2012; Hugdahl et al., 2006; Schöning et al., 2007; Weiss et al., 2003b). Perhaps with larger samples a clearer distinction between male-biased activity in the parietal lobe versus female-biased activity in the frontal lobe may become apparent. For now, however, the overarching result is a lack of consistent s/g differences, likely due to low statistical power and false positives (David et al., 2018).

The same conclusion applies to the remainder of fMRI studies that used other spatial cognition tasks (Table 9C). After MRT, the next most common spatial task for assessing male-female difference is virtual navigation. At the behavioral level, males outperform females on navigation and wayfinding tasks by a pooled effect size of $d = 0.36$ in adults and $d = 0.15$ in children (Nazareth et al., 2019). We found four studies that compared fMRI activation between males and females during virtual navigation, which involves finding one's way through a featureless 3-dimensional computer maze. The first and most highly-cited study found that males exhibited stronger activation of the left hippocampus, whereas females exhibited stronger right prefrontal and left parietal activation during virtual maze exploration (Grön et al., 2000). However, a larger study by Ohnishi et al. (2006) found no s/g difference in hippocampal activation, but did detect stronger activation of the parietal cortex by poor navigators, both male and female. Another study reported greater activity in the right hippocampus of males relative to females (Persson et al., 2013), the opposite side from Grön et al. (2000). Finally, Sneider et al. (2011) observed s/g differences in right frontal and left temporal gyri, but similar bilateral hippocampal activation in both sexes.

Taken together, these studies reveal no consistent s/g differences in regional brain activation during spatial navigation, but rather, considerable similarity in activation of visual and parahippocampal structures between males and females. Such similarity is further supported by a more recent assessment of the spatial navigation network derived from rsfMRI connection analysis. In this much larger study, no s/g differences were identified in the connections between three key areas known to participate in spatial navigation: hippocampus, parahippocampal place area, and retrosplenial areas of the parietal lobe (Boccia et al., 2017). In summary, existing research has not elucidated a neural basis for the male/female difference in spatial navigation, much like collective fMRI studies of MRT.

In the interest of completeness, Table 9C includes 11 other fMRI studies comparing men and women during spatial task performance. Three of these assessed line angle judgement, which is notable since this task produced the largest performance difference ($d = 0.57$ favoring males) among several cognitive tests in a massive internet survey (Collaer et al., 2007). Of the three fMRI studies, all detected at least one area of greater activation in males, but none of these replicated between studies (Clements et al., 2006; Clements-Stephens et al., 2009; Gur et al., 2000). Among the remaining eight spatial task studies, five reported stronger activation of the precuneus in males compared to females. Males' precuneus was also more strongly activated in 4 of the 19 MRT studies (Carrillo et al., 2010; Semrud-Clikeman et al., 2012). Beyond this minor agreement, however, the overarching conclusion is that spatial performance is not reliably associated with any specific male/female difference in brain activation.

Somewhat different conclusions were reached in a recent meta-

analysis by Yuan et al. (2019). These authors divided such skills into large-scale, such as navigation, versus small-scale such as mental rotation. Overall, they found largely similar activation in males and females across the brain, but did detect some areas of stronger activation in females: parahippocampal gyrus and lentiform nucleus during large-scale processing, and right sub-gyral and precuneus plus left middle frontal gyrus during small-scale processing. These results do not agree with any of the findings in Tables 9A, 9B, 9C, probably because Yuan et al. (2019) separately pooled all-male and all-female activation data, as opposed to pooling the differences between male and female fMRI activations within the same experiments. It is also difficult to reconcile the stronger brain activation in females with their weaker average spatial performance, although Yuan et al. (2019) conclude, rather speculatively, that it is because women are "more susceptible to emotions."

Of course, s/g differences in spatial skills must be mediated somehow by brain-level differences, but even assuming subtle neural differences are someday agreed upon, none of this research addresses their origin. The gradual emergence of spatial performance gaps over childhood (Lauer et al., 2019), together with the extreme plasticity of hippocampal circuits (Erickson et al., 2011; Hufner et al., 2011; Woollett and Maguire, 2011; Fotuhi et al., 2012; Draganski et al., 2006; Noble et al., 2012; Luders et al., 2013; Suo et al., 2012) suggests that s/g differences are largely acquired through practice. Indeed, research addressing mental rotation ability in men and women with similar technical experience—that is, trained military pilots—found no s/g difference in either mental rotation (Verde et al., 2013) or navigational (Verde et al., 2015) abilities, consistent with a lack of inborn "sexual dimorphism" in spatial cognitive circuitry (see also Jones et al., 2003; Maguire et al., 1999).

7.2. Emotion processing

A last behavioral domain in which fMRI has been extensively used to search for male/female differences is emotion processing. Indeed, so many experiments have been conducted to address this issue that we can limit this review to the findings from seven separate meta-analyses of fMRI activations in response to emotionally evocative stimuli (Table 10).

Like verbal and spatial skills, s/g differences in emotional expression and affective personality traits have been widely documented at the behavioral level. In this domain, the effect sizes tend to be small-to-moderate (Brody and Hall, 2008; Else-Quest et al., 2012) and smaller still in childhood (Chaplin and Aldao, 2013). Much of this research has focused on empathy, or the ability to detect the emotional experience of another person. Empathy can be assessed subjectively, based on participants' self-report of caring or feeling in response to various social scenarios. Such subjective self-assessment tends to produce a large female advantage ($d \approx 1.0$). By contrast, objective measures of empathy, such as participants' accuracy at identifying the emotion expressed in photos or video clips of others' faces (Hall, 1978), are more modest. According to a meta-analysis of 215 independent samples (Thompson and Voyer, 2014), females outperform males at identifying emotion in non-verbal displays by an effect size of only $d = 0.19$. Again, the effect is smaller in children ($d = 0.13$), according to meta-analysis by McClure (2000).

Such small behavioral effects may explain why fMRI researchers have had difficulty finding consistent brain s/g differences during objective emotion tasks. Three early studies claimed that a lateralization difference underlies a s/g difference in emotional memory (Cahill et al., 2001, 2004; Canli et al., 2002). These were summarized in a review by Hamann (2005) that stated emotional memories are predominantly encoded by the right amygdala in males but the left amygdala in females. However, these lateralization differences have not held up in the several meta-analyses of brain emotion processing.

The first meta-analysis to assess s/g difference in emotion processing was by Wager et al. (2003) who compared 22 studies of men to 14

Table 10

Meta-analyses of sex/gender difference in emotion processing based on fMRI activation.

Meta-analysis	# Studies	# M/F	Sex/gender difference findings
Wager et al. (2003)	65	n.r.	<ul style="list-style-type: none"> No differences in whole brain activation Same lateralization pattern but stronger in males (R. temporal cortex) F>M activation in brainstem and cerebellum
Sergerie et al. (2008)	148	1075 / 1146	<ul style="list-style-type: none"> No differences in lateralized amygdala activity M>F amygdala activation; correlation between effect size and proportion M subjects
Lamm et al. (2011)	32	n.r.	<ul style="list-style-type: none"> No differences in brain activation during empathy for pain
Stevens & Hamann (2012)	35 M-only; 37 F-only; 7 both M&F	561 / 656	<ul style="list-style-type: none"> Negative emotion: F>M in left amygdala and hippocampus. M>F in right precentral gyrus, inferior frontal gyrus, and insula Positive emotion: M>F in left amygdala No difference in amygdala activation when all studies analyzed regardless of valence
Sacher et al. (2013)	18	271 / 264	<ul style="list-style-type: none"> M>F activation in bilateral amygdala and left orbitofrontal cortex F>M activation in several clusters including cingulate cortex, globus pallidus, putamen, thalamus & midbrain
Garcia-Garcia et al. (2016)	12	148/150	<ul style="list-style-type: none"> None (negative emotional stimuli activate right amygdala in both M and F)
Filkowski et al. (2017)	56	941/961	<ul style="list-style-type: none"> M>F mPFC/ACC, frontal pole, mediodorsal nucleus of thalamus F>M bilateral amygdala, hippocampus, regions of the dorsal midbrain

Sex/gender similar findings show in green; areas of higher activation in males shown in blue; areas of higher activation in females shown in pink. Emphasis is on amygdala findings but other key findings are also included. #M/F = total number of male and female participants across studies in each meta-analysis.

studies of women and found no difference in brain areas activated in response to emotional stimuli, including the amygdala, although males were more lateralized in one region (temporal cortex). They also noted a greater degree of activation of the brainstem and cerebellum in studies of women, compared to studies of men.

The next meta-analysis was by Sergerie et al. (2008) who specifically focused on the amygdala. Their analysis included 42 studies and contrary to the conclusion of Hamann (2005), found no s/g difference in lateralized amygdala activity. Sergerie et al. (2008) did, however, did find stronger amygdala activation overall in men in response to emotional stimuli. Enhanced male bilateral amygdala activation was similarly detected in a smaller meta-analysis by Sacher et al. (2013), who also reported greater male activation in the left orbitofrontal cortex and female activation in several clusters including the cingulate cortex, globus pallidus, putamen, thalamus and midbrain.

Around this time, Lamm et al. (2011) conducted a meta-analysis of brain activation focused on participants' experience of empathy while witnessing others in pain. Although s/g differences in brain activation were expected based on well-described behavioral differences (Hall, 1978; Thompson and Voyer, 2014), this synthesis of 32 studies detected none. Lamm et al. (2011) found that empathy for pain activates the bilateral anterior insula and anterior-medial cingulate gyrus equally in men and women.

Stevens and Hamann (2012) took a different approach in their meta-analysis of 79 studies, dividing experiments according to the valence of emotional stimuli presented in the scanner. They found that negative emotional stimuli produced greater activation in females' left amygdala and hippocampus and in males' right precentral gyrus, inferior frontal gyrus, and insula. However, for positive emotion, the activation difference in left amygdala was reversed, appearing higher in males. When all stimuli were analyzed together, regardless of valence, there was no difference in amygdala activation between men and women.

Garcia-Garcia et al. (2016) followed this valence approach but were unable to replicate the s/g findings of Stevens and Hamann (2012). Rather, they found that both males and females show stronger activation of left over right amygdala in response to negative-valenced stimuli. They attribute the difference from Stevens and Hamann (2012) to their

stricter inclusion criteria, which disallowed PET studies, very small studies (<12 participants), and studies in which males and females were not matched for age. Although this reduced their total sample to 12 studies, Garcia-Garcia et al. (2016) concluded that this more homogeneous dataset demonstrated "robust similarities" between men and women in brain emotion processing.

Finally, the most recent meta-analysis by Filkowski et al. (2017) included 56 studies and found results inconsistent with all six prior meta-analyses. Thus, they reported greater bilateral amygdala and hippocampal activity in females in response to visual emotional stimuli of any valence. This is opposite the result in Sacher et al. (2013) and Sergerie et al. (2008). Filkowski et al. (2017) did, however, replicate the finding of Wager et al. (2003) that women showed stronger brainstem activation in response to emotional stimuli. Males, by contrast, showed stronger activation in several frontal areas, a result that agrees better with Sacher et al. (2013) and Stevens and Hamann (2012), albeit the latter found this only for negative emotional stimuli.

In summary, fMRI has been extensively utilized over the past quarter century to compare the neural basis of emotion processing between males and females, with surprisingly little consensus. Hypotheses concerning lateralization differences or valence-specific s/g activity patterns have not held up across meta-analyses, much less many individual studies. Technical factors likely contribute to the lack of reproducibility in s/g effects (Botvinik-Nezer et al., 2020): differences in MRI hardware, image processing pipeline, spatial data analysis (Hong et al., 2019), stimulus type, age and nationality of subjects, educational background, and especially, sample size (Button et al., 2013; Szucs and Ioannidis, 2017). Another limitation is the variability in behavioral performance during scanning, which was not addressed in any of the meta-analyses in Table 10. We did track this for verbal (Tables 8A, 8B, 8C, 8D) and spatial (Tables 9A, 9B, 9C) fMRI studies and found only a minority observed the expected s/g difference in task accuracy or reaction time in the scanner. This can be taken as further evidence of under-powering (for studies that fail to find a behavioral difference). Conversely, in studies that did find a performance difference between males and females, any fMRI signal differences could reflect the degree of challenge more than cognitive circuitry per se, complicating the interpretation of such neural findings.

If anything can be concluded from these hundreds of fMRI studies of

emotion processing, it is that men and women both activate these same neural structures: amygdala, hippocampus, thalamus, frontal lobe, and midbrain. In other words, the circuitry for emotion processing—like language and spatial cognition—appears overwhelmingly similar between men and women. Although we cannot eliminate the possibility that there are reliable male/female differences in certain populations under very specific conditions, the parsimonious conclusion is that group-level s/g differences, if they exist, are buried within the greater variance of individual difference. Very similar conclusions were reached about s/g difference in a recent meta-analysis of fMRI responses during sexual arousal (Mitricheva et al., 2019).

8. Discussion

As this comprehensive review has demonstrated, s/g difference in human brain structure and function has been extensively analyzed over the past several decades. The extent of this evidence belies the plaint that women's brains have been overlooked in neuroscience research (e.g., Ferretti et al., 2018; Woodruff et al., 2014). It is true that many studies do not disaggregate male and female data, opting to treat s/g as a nuisance variable (Barnes et al., 2010; Yan et al., 2013). Nonetheless, the data have been consistently collected and are widely available. As we've seen, women and men have been equally included in neuroimaging research since the dawn of brain MRI, an admirable achievement given the importance of diverse participation for generating scientific advances (Clayton, 2016).

Indeed, one reason there has been so much research on s/g difference is simple convenience. The fact of equal inclusion means that neuroimaging researchers can take the opportunity to publish a novel paper by simply dividing up their existing participant pool and doing a subgroup analysis for male/female difference. So while a case can be made for requiring greater disaggregation of male/female data in future publication (Arnegard et al., 2020), the existing brain MRI literature is sufficiently large to allow us to answer many questions about human s/g brain difference. And for most of these questions, the absence of difference is not due to an absence of data.

Another reason why this topic has been so extensively studied is the considerable extra-scientific interest in it. Given the potential behavioral implications of brain s/g difference and the opportunity to link such differences to an array of medical and societal gender gaps—e.g., neurologic and behavioral health, personality, relationships, education, occupations, leadership—any difference found between male and female brains attracts considerable attention from journal editors, university press officers, journalists, social media, and lay readers (Eliot, 2011; Fine, 2010; Maney, 2015). This desire to explain social gender gaps through brain data may also explain why older, underpowered and often unreplicated studies continue to be found and widely cited (e.g., Cahill et al., 2001; DeLacoste-Utamsing and Holloway, 1982; Filipek et al., 1994; Goldstein et al., 2001; Grön et al., 2000; Shaywitz et al., 1995). Another potent incentive has emerged from funding opportunities, since the NIH and other agencies have increasingly prioritized male-female comparison across preclinical and clinical studies (Clayton, 2016). Considering all of these motives, it is not surprising that neuroscience has produced the very large stockpile of data on s/g effects that we have attempted to synthesize in this lengthy review.

But for all the surplus of brain-level data on male-female difference, surprisingly few clear findings have emerged, and even less to justify labeling the human brain as "sexually-dimorphic." Nor does anything in this massive data collection actually explain male/female differences in psychology or mental health (De Vries and Södersten, 2009; Hirnstein et al., 2019) in spite of decades of such promise. To the contrary, the data show that male and female brains are overwhelmingly similar, or monomorphic, and suggest that finding such neural correlates will more fruitfully be achieved through study at the individual, as opposed to s/g group level.

8.1. Brain size and its correlates

What we do know is that brain size is reliably larger in males. This difference holds throughout the lifespan and parallels s/g difference in overall bodily size and growth. Thus, the male-female difference in TBV increases from 6% at birth to 11% in adulthood, while the difference in average body mass increases from 4% to 18% over the same period. This brain-body parallel is also evident in the pace of growth, as females reach maximum brain volume about 2 years earlier than males (age 10.5 vs. 12.5, Lenroot et al., 2007; but see Wierenga et al., 2014). This gap is similar to females' earlier onset of puberty and attainment of adult height, although interestingly, brain volume maximization appears to lead the pubertal height spurt by 1.5 years in both sexes (Tanner, 1990, p. 60).

With the difference in overall brain size comes other male/female brain differences that are largely, if not exclusively, attributable to size rather than sex. One of these is GM/WM ratio, which averages 5.5% larger in females across multiple studies. As brain size increases, there is a disproportionate increase in the denominator of this ratio, since larger brains need larger-caliber, more heavily myelinated axons to transmit action potentials across greater distances (Bush and Allman, 2003; Zhang and Sejnowski, 2000). Thus, the s/g difference in GM/WM ratio is largely eliminated when adjusted for total brain size (Leonard et al., 2008; Luders et al., 2002; Jäncke et al., 2015).

Without exception, the difference in global brain size is reflected in every major cortical and subcortical brain component, which range from 5 to 11% larger in absolute volume in males compared to females (Ritchie et al., 2018; Voevodskaya et al., 2014). Again, normalizing these measures to individual brain or head size largely eliminates any volume difference between males and females in specific structures. Of the handful of s/g differences that survive this correction, all are small in magnitude ($d \leq 0.25$) and sensitive to analytic methods—brain segmentation software or brain size normalization method (Table 3). In an effort to skirt this problem, at least four studies (Table 2) have compared male and female structures in participant pools matched for brain size, but no s/g differences have replicated between them.

Among studies based on large, diverse samples ($N > 2,000$; Table 1), small differences ($d \approx 0.18$) in relative volumes have been identified fairly consistently for two subcortical structures: the amygdala and putamen (Lotze et al., 2019; Potvin et al., 2016, 2018a; Ritchie et al., 2018). However, the absolute difference in amygdala volume is a mere 1% larger in males (after ICV or TBV normalization) and not reliably significant (Marwha et al., 2017). Moreover, the magnitude and significance of s/g difference in both of these structures is highly sensitive to the precise method of brain volume normalization (e.g., Coupé et al., 2017; Fjell et al., 2009; Pintzka et al., 2015).

For cortical volumes there is even less consistency in claims of s/g difference (Table 3), especially comparing studies using surface- versus volume-based segmentation. Two large Freesurfer studies (Potvin et al., 2017, 2018a; Ritchie et al., 2018) did agree on a handful of structures, but these were not replicated in large VBM studies. Moreover, the functional significance of the Freesurfer findings is difficult to parse, since women outperform men in verbal and social-emotional skills (Hyde and Linn, 1988; Zell et al., 2015), which rely on areas found to be larger in men in these studies (e.g., fusiform, orbitofrontal and pars triangularis). Also paradoxically, men outperform women in physical and spatial skills, which depend on areas found to be larger in women (superior parietal, postcentral and paracentral gyri) in these large studies. So these findings fail to support the claim that the male human brain is evolutionary augmented for spatial-mechanical cognition (so-called "folk physics") and the female brain for language and social cognition ("folk psychology") as posited by Geary (2019) based on a few smaller and older studies.

Cortical thickness is another metric often stated to differ between males and females. Several large and small MRI studies report greater average cortical thickness in females (Table 4), however, this finding is

highly sensitive to image acquisition method (Govindarajan et al., 2014). Moreover, these MRI findings contradict results from more precise histological measurements, which trend towards greater thickness in males. Unlike structural volumes, cortical thickness does not vary with overall brain or intracranial volume (Potvin et al., 2017, 2018a); hence it is not appropriate to normalize to 3-dimensional volume when comparing males and females in this 1-dimensional measure, and studies that have done so overestimate cortical thickness in females (Im et al., 2008).

Thus, once we account for individual differences in brain size, there is almost no difference in the volume of specific cortical or subcortical structures between men and women. One tiny exception is INAH-3, which is too small (0.1 mm) to be visible by MRI but has been independently confirmed by four post-mortem histological studies to be about 1.6-fold larger in men (Allen et al., 1989; Byne et al., 2000; Garcia-Falgueras and Swaab, 2008; LeVay, 1991). Although discovered by analogy to the rodent SDN-POA and speculated to participate in reproductive behavior, neither INAH-3 nor the SDN-POA has been assigned a clear behavioral function (McCarthy, 2016).

8.2. Minute differences in lateralization and interhemispheric connectivity

Along with regional brain volume differences, much of the research on human s/g brain difference has focused on asymmetry. A long-standing claim is that men's brains are less efficiently connected between hemispheres and therefore operate in a more lateralized fashion than women's. Early on, findings from the clinical aphasia literature were cited to support this claim (McGlone, 1980), but this has not held up to more recent analyses (Plowman et al., 2012; Watila and Balarabe, 2015). Much more research has been conducted to test this hypothesis using both structural and functional imaging. On the structural side, most studies have focused on the corpus callosum (Table 5) which, like all brain structures, is larger in males in raw volume or cross-sectional area. However, when such callosal measures are normalized by overall brain size, this difference usually disappears (e.g., Jäncke et al., 2015; Luders et al., 2014; Voevodskaya et al., 2014) or ends up slightly larger in females. Importantly, the difference is very small, with sex accounting for less than 1% of the population variance in callosal volume (Potvin et al., 2016, 2018a). Similarly, the anterior commissure has been repeatedly analyzed for s/g difference (Table 6a), producing inconsistent results that add up to very weak evidence that the structure is augmented in females. Evidence is more consistent that the massa intermedia (or, interthalamic adhesion) is larger and more often present in female than in male brains (Table 6b). However, it is not known whether this is a true s/g difference or a function of brain size, given the mechanical stress on this small bridge of non-neuronal tissue spanning the third ventricle.

On the functional side, the evidence is mixed supporting slightly greater interhemispheric connectivity in females or conversely, greater lateralization in males. Exhaustive analyses of dichotic behavioral tasks have provided evidence for very weak, population-level enhancement of lateralization in males across auditory, visual, tactile and dual-task processing (Hiscock et al., 1994, 1995, 1999, 2001), with s/g accounting for just 0.1 % of the population variance (Voyer, 1996). Among fMRI studies of language lateralization, meta-analysis found no s/g difference (Sommer et al., 2008), whereas fMRI studies comparing resting connectivity between males and females have produced inconsistent results with respect to lateralization and asymmetry, even among studies using the same 1000 Functional Connectomes database (see Section 5). Greater attention has been paid to connectivity estimated from structural covariance measures. The most-cited such study claims that interhemispheric connectivity is relatively stronger in females, and intrahemispheric stronger in males (Ingallalikar et al., 2014). However, that finding has been challenged for ignoring the influence of brain size, which can fully account for the relative density of inter- versus intra-hemispheric connections independent of s/g (Dennis et al., 2013;

Hänggi et al., 2014; Martinez et al., 2017). Overall, the weak and inconsistent differences in male/female interhemispheric connectivity explain why efforts to link lateralization to cognitive sex differences have been unsuccessful (Hirnstein et al., 2019).

8.3. Connectome differences and sex/gender prediction

In spite of their ambiguous conclusions about lateralization, the explosion of new approaches for estimating brain connectivity has been a boon for s/g difference research. The large sample sizes and massive amount of data collected across various collaborative imaging initiatives have spawned many dozens (Suppl. Table 3) of attempts to compare connectivity architecture between male and female brains. Nonetheless, the literature on s/g connectome difference is starkly conflicted and inconclusive. Although some consensus can be found indicating higher efficiency and DMN connectivity in females, there is dispute about the degree of integration across networks in each sex and once again, the confounding influence of brain size has been inconsistently examined.

The one place where large-scale comparison of males and female brains has been highly successful is for sex prediction based on machine learning (Table 7). With so many variables and such large samples, artificial intelligence is very good (~90 % accuracy) at predicting whether a particular scanned brain came from a male or female participant. However, this predictive validity is highly diminished or even eliminated when brain size is factored out of the analysis (Chekroud et al., 2016; Sanchis-Segura et al., 2020; Zhang et al., 2018). Moreover, the criteria used to discriminate male and female brains in one training sample have not performed well for s/g discrimination in ethnically or geographically distinct samples (Joel et al., 2018; Yang et al., 2020a). It is likely that differences in head shape (Bijsterbosch et al., 2018; Zilles et al., 2001) between males and females influence these blind multivariate discrimination analyses. In spite of endorsing a “dimorphic” sorting of male and female brains, this line of research is only able to predict in one direction—that is, whether a particular subject is male or female, based largely on brain size; it is not capable of advancing its deeper aim—predicting an individual's specific brain attributes based on knowledge of their sex or gender.

8.4. Task-based fMRI has not elucidated male-female differences in cognitive neural networks

If any approach seemed likely to reveal reliable s/g neural differences, it is task-based fMRI. Male-female differences in personality, interests, and cognitive and emotional abilities have been well-validated at the psychological level and must be mediated at some level by brain structure and function. Assays for brain activity during task performance routinely subtract resting activity, narrowing the focus to circuits underlying the precise behavior known to differ between males and females. Moreover, such performance-based activity is likely to be strongly shaped by subjects' prior practice and learning, which tends to magnify male-female difference over the lifespan, thereby amplifying potential brain differences.

We surveyed three cognitive domains with the best characterized behavioral s/g differences but found no clearly defined neural differences across these many dozens of studies. In the language domain, differential lateralization has been disproven (Sommer et al., 2008), and there has been virtually no replication of male/female differences across 33 studies assessing regional activation, even after partitioning into verbal fluency, verb generation, and verbal memory tasks (Tables 8A, 8B, 8C, 8D). We concur with other reviewers that there is no reliable evidence for s/g difference in neural language processing in either adults (Ihnen et al., 2009; Kaiser et al., 2009; Wallentin, 2009) or children (Etchell et al., 2018). In the domain of spatial processing, at least 39 fMRI studies have reported a wide range of findings in males vs. females. Although many have detected structural, functional, or lateralization differences among key nodes for spatial cognition—including the

hippocampus, parahippocampal place area, retrosplenial cortex, intraparietal sulcus, and medial superior premotor area—there has been almost no replication, even among studies using identical spatial tasks (mental rotation, virtual navigation, or judgement of line orientation; Tables 9A, 9B, 9C). And finally, in the domain of emotion processing, seven existing meta-analyses have failed to reach consensus about s/g difference in amygdala and other limbic structure activation, but instead demonstrate overwhelming male-female similarity in the circuits for processing empathy and other emotional experiences (Table 10).

There are limitations to our analysis, which was not fully systematic although we relied on prior meta-analyses whenever possible. In particular, we utilized the systematic review and meta-analysis by David et al. (2018) as a backstop for identifying fMRI studies of s/g effects between 2004–2013. Given the lack of replication within cognitive domains and the finding by David et al. (2018) that larger studies were no more likely to report significant s/g differences than smaller studies, we concur that the literature in this area is contaminated by many false positives. This problem is worse for fMRI than for structural MRI studies of s/g difference since sample sizes are typically 10 or more times smaller in functional studies.

From another perspective, our inability to find reliable s/g difference in task-based fMRI activity is entirely unsurprising (Cosgrove et al., 2007). Such studies are now appreciated to be massively underpowered (Button et al., 2013) and to exhibit such broad flexibility in data collection and analytic pipelines as to virtually guarantee high rates of false positives (Carp, 2012) and poor reproducibility (Botvinik-Nezer et al., 2020). Small sample size is largely responsible for the poor reliability of task-based fMRI studies documented by Elliott et al. (2020); based on meta-analysis of 90 experiments, they report a mean intraclass correlation of just 0.397 for test-retest reliability and conclude that a sample size of 217 participants is the minimum required to achieve even moderate reliability. This sample size is 5–10 times greater than those typically studied in any of the task-based fMRI analyses of s/g difference we collected (Tables 8A, 8B, 8C, 8D–10). Notably, Elliott et al. (2020) conclude that task-based fMRI studies are not currently suitable for brain biomarker discovery or any type of research on individual differences, which of course includes s/g difference.

8.5. The human brain is not sexually dimorphic

Summarizing across the extensive findings we reviewed, s/g differences in the human brain are extremely subtle and variable. There is nothing to justify the term “sexual dimorphism” to describe them. Among the few reliable differences, nearly all are byproducts of brain size, and none are evidence of “two shapes” as “dimorphism” would denote. Thus, when brain size is covaried in the analysis of individuals’ brain measures, s/g explains about 1% of the total variance. In other words, brain differences attributable to sex or gender are trivial relative to other sources of individual variation. Furthermore, differences that are often portrayed as related to s/g (e.g., GM/WM ratio, or inter- vs. intrahemispheric connectivity ratio) are more accurately attributable to brain size, such that they distinguish large- from small-headed men (or large- from small-headed women) as well as they distinguish the average man from the average woman.

Indiscriminate use of “sexual dimorphism” has been rejected by leaders in the field (Ball et al., 2014; McCarthy et al., 2012), who state that it should be restricted to phenomena in which there is no overlap between groups, such as the shape of ovaries vs. testes or SNB volume in male vs. female rodents. For smaller effects, these authors advise using “sex difference” or “sex effect.” Some in the neuroimaging field are also starting to abandon “sexual dimorphism” (Gur and Gur, 2016; Kaczkurkin et al., 2019; but see Anderson et al., 2019). We agree and argue that the issue is more than semantic. The term “dimorphism” has potent heuristic value, reinforcing the belief in categorically distinct organs: a “male brain” and “female brain” that have been evolutionarily shaped to produce two psychologically distinct categories of people (c.f.,

Brizendine, 2006, 2010; Goldman, 2017; Murray, 2020).

In other words, the term “dimorphism” reinforces a binary understanding of s/g brain difference, when in fact, few such differences actually exist and the ones that do are very small, with great variability from population to population. Of note, a similar conclusion is converging from the growing research on transgender participants’ brains, according to Smith et al. (2015) who remark that “viewing gender as a binary or dichotomous category has to be reconsidered.”

Rather, a picture is emerging not of two brain types nor even a continuous gradient from masculine to feminine, but of a multidimensional “mosaic” of countless brain attributes that differ in unique patterns across all individuals (Joel et al., 2015). Although such differences may, in a particular sample, sum up to discriminate male from female brains, the precise discriminators do not translate across populations (Table 7; see also Joel et al., 2018; Sanchis-Segura et al., 2020) so are not diagnostic of two species-wide types. In this sense, the brains of male and females are not dimorphic (like the gonads) but monomorphic, like the kidneys, heart and lungs, which can be transplanted between women and men with great success.

9. Conclusion

Scholarly interest in brain sex difference is as old as Aristotle (Deslauriers, 2009). Despite clear behavioral differences between men and women, s/g differences in the brain are small and inconsistent, once individual brain size is accounted for. Most neuroscientists assume this ambiguity will be solved through technical improvements: that larger studies, using higher resolution imaging and better processing pipelines will uncover the “real,” or species-wide differences between male and female brain structure and connectivity patterns. However, the present synthesis indicates that such “real” or universal sex-related difference do not exist. Or at best, they are so small as to be buried under other sources of individual variance arising from countless genetic, epigenetic, and experiential factors. Thus, s/g differences in brain architecture may be similar to sex effects in gene-phenotype architecture; while statistically discernable in a very large (>100,000) sample, such effects contributed only 1.4 % to the accuracy of genotype-phenotype prediction (Rawlik et al., 2016).

In layperson’s terms, these findings can be interpreted as rebutting popular discourse about the “male brain” and “female brain” as distinct organs. They also have relevance to research on the many neurobehavioral disorders that differ in prevalence between men and women, such as autism, ADHD, dyslexia, depression, anxiety, dementia and eating disorders. Although studies of s/g brain difference often begin from the premise that they will lead to a better understanding of these health disparities, their actual data seem unlikely to advance the aims of precision medicine.

Acknowledgements

This work was supported by a grant from the Fred B. Snite Foundation.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.neubiorev.2021.02.026>.

References

- Achiron, R., Lipitz, S., Achiron, A., 2001. Sex-related differences in the development of the human fetal corpus callosum: *in utero* ultrasonographic study. *Prenat. Diagn.* 21, 116–120.
- Agcaoglu, O., Miller, R., Mayer, A.R., Hugdahl, K., Calhoun, V.D., 2015. Lateralization of resting state networks and relationship to age and gender. *Neuroimage* 104, 310–325.

- Allen, L.S., Gorski, R.A., 1991. Sexual dimorphism of the anterior commissure and massa intermedia of the human brain. *J. Comp. Neurol.* 312, 97–104.
- Allen, L.S., Gorski, R.A., 1992. Sexual orientation and the size of the anterior commissure in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 89, 7199–7202.
- Allen, L.S., Hines, M., Shryne, J.E., Gorski, R.A., 1989. Two sexually dimorphic cell groups in the human brain. *J. Neurosci.* 9, 497–506.
- Allen, E.A., et al., 2011. A baseline for the multivariate comparison of resting state networks. *Front. Syst. Neurosci.* 5 (2011).
- Allendorfer, J.B., Lindsell, C.J., Siegel, M., Banks, C.L., Vannest, J., Holland, S.K., Szaflarski, J.P., 2012. Females and males are highly similar in language performance and cortical activation patterns during verb generation. *Cortex* 48, 1218–1233.
- Anderson, N.E., Harenski, K.A., Harenski, C.L., Koenigs, M.R., Decety, J., Calhoun, V.D., Kiehl, K.A., 2019. Machine learning of brain gray matter differentiates sex in a large forensic sample. *Hum. Brain Mapp.* 40, 1–11.
- Andreano, J.M., Cahill, L., 2009. Sex influences on the neurobiology of learning and memory. *Learn. Mem.* 16, 248–266.
- Ardekanli, B.A., Figarsky, K., Sidtis, J.J., 2013. Sexual dimorphism in the human corpus callosum: an MRI study using the OASIS brain database. *Cereb. Cortex* 23, 2514–2520.
- Arnegard, M.E., Whitten, L.A., Hunter, C., Clayton, J.A., 2020. Sex as a biological variable: a 5-year progress report and call to action. *J. Wom. Health* 29, 858–864.
- Arnold, A.P., 2004. Sex chromosomes and brain gender. *Nat. Rev. Neurosci.* 5, 701.
- Arnold, A.P., Gorski, R.A., 1984. Gonadal steroid induction of structural sex differences in the central nervous system. *Ann. Rev. Neurosci.* 7, 413–442.
- Azim, E., Mobbs, D., Jo, B., Menon, V., Reiss, A.L., 2005. Sex differences in brain activation elicited by humor. *Proc. Natl. Acad. Sci. U. S. A.* 102, 16496–16501.
- Ball, G.F., 2016. Species variation in the degree of sex differences in brain and behaviour related to birdsong: adaptations and constraints. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150117.
- Ball, G.F., Balthazard, J., McCarthy, M.M., 2014. Is it useful to view the brain as a secondary sexual characteristic? *Neurosci. Biobehav. Rev.* 46, 628–638.
- Barnes, J., Ridgway, G.R., Bartlett, J., Henley, S.M., Lehmann, M., Hobbs, N., Clarkson, M.J., MacManus, D.G., Ourselin, S., Fox, N.C., 2010. Head size, age and gender adjustment in MRI studies: a necessary nuisance? *Neuroimage* 53, 1244–1255.
- Beacher, Felix D.C.C., Radulescu, E., Minati, L., Baron-Cohen, S., Lombardo, M.V., Lai, M., Walker, A., Howard, D., Gray, M.A., Harrison, N.A., Critchley, H.D., 2012. Sex differences and autism: brain function during verbal fluency and mental rotation. *PLoS One* 7, e38355.
- Berenbaum, S.A., Denburg, S.D., 1995. Evaluating the empirical support for the role of testosterone in the Geschwind-Behan-Galaburda model of cerebral lateralization—commentary. *Brain Cogn.* 27, 79–83.
- Bijsterbosch, J.D., Woolrich, M.W., Glasser, M.F., Robinson, E.C., Beckmann, C.F., Van Essen, D.C., Harrison, S.J., Smith, S.M., 2018. The relationship between spatial configuration and functional connectivity of brain regions. *eLife* 7, e32992.
- Bishop, K.M., Wahlsten, D., 1997. Sex differences in the human corpus callosum: Myth or reality? *Neurosci. Biobehav. Rev.* 21, 581–601.
- Biswal, B.B., et al., 2010. Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U. S. A.* 107, 4734–4739.
- Björnholm, L., Nikkinen, J., Kiviniemi, V., Nordström, T., Niemelä, S., Drakesmith, M., Evans, J.C., Pike, G.B., Veijola, J., Paus, T., 2017. Structural properties of the human corpus callosum: multimodal assessment and sex differences. *Neuroimage* 152, 108–118.
- Bluhm, R.L., Osuch, E.A., Lanius, R.A., Boksman, K., Neufeld, R.W.J., Théberge, J., Williamson, P., 2008. Default mode network connectivity: effects of age, sex, and analytic approach. *Neuroreport* 19, 887–891.
- Boccia, M., Sulzio, V., Nemmi, F., Guariglia, C., Galati, G., 2017. Direct and indirect parieto-medial temporal pathways for spatial navigation in humans: evidence from resting-state functional connectivity. *Brain Struct. Funct.* 222, 1945–1957.
- Boghi, A., Rasetti, R., Avidano, F., Manzone, C., Orsi, L., D'Agata, F., Caroppo, P., Bergui, M., Rocca, P., Pulvirenti, L., Bradac, G.B., Bogetto, F., Mutani, R., Mortara, P., 2006. The effect of gender on planning: an fMRI study using the Tower of London task. *Neuroimage* 33, 999–1010.
- Boles, D., 2005. A large-sample study of sex differences in functional cerebral lateralization. *J. Clin. Exp. Neuropsychol.* 27, 759–768.
- Botvinick-Nezer, R., Holzmeister, F., Camerer, C.F., Dreber, A., Huber, J., Johannesson, M., Kirchner, M., Iwanir, R., Mumford, J.A., Adcock, R.A., Avesani, P., 2020. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582, 84–88.
- Brain Development Cooperative Group, 2012. Total and regional brain volumes in a population-based normative sample from 4 to 18 years: the NIH MRI study of normal brain development. *Cereb. Cortex* 22, 1–12.
- Breedlove, S.M., Arnold, A.P., 1980. Hormone accumulation in a sexually dimorphic motor nucleus of the rat spinal cord. *Science* 210, 564–566.
- Brizendine, L., 2006. The Female Brain. Morgan Road Books., New York.
- Brizendine, L., 2010. The Male Brain. Three Rivers, New York.
- Brody, L.R., Hall, J.A., 2008. Gender and emotion in context. In: Lewis, M., Haviland-Jones, J.M., Barrett, L.F. (Eds.), *Handbook of Emotions*. Guilford Press, New York, pp. 395–408.
- Bryden, M.P., McManus, I.C., Bulman-Fleming, M., 1994. Evaluating the empirical support for the geschwind-behan-galaburda model of cerebral lateralization. *Brain Cogn.* 26, 103–167.
- Burman, D.D., Bitan, T., Booth, J.R., 2008. Sex differences in neural processing of language among children. *Neuropsychologia* 46, 1349–1362.
- Bush, E.C., Allman, J.M., 2003. The scaling of white matter to gray matter in cerebellum and neocortex. *Brain Behav. Evol.* 61, 1–5.
- Butler, T., Imperato-McGinley, J., Pan, H., Voyer, D., Cordero, J., Zhu, Y., Stern, E., Silbersweig, D., 2006. Sex differences in mental rotation: top-down versus bottom-up processing. *Neuroimage* 32, 445–456.
- Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365.
- Byne, W., Lasco, M.S., Kemether, E., Shinwari, A., Edgar, M.A., Morgello, S., Jones, L.B., Tobet, S., 2000. The interstitial nuclei of the human anterior hypothalamus: an investigation of sexual variation in volume and cell size, number and density. *Brain Res.* 856, 254–258.
- Byne, W., Tobet, S., Mattiace, L.A., Lasco, M.S., Kemether, E., Edgar, M.A., Morgello, S., Buchsbaum, M.S., Jones, L.B., 2001. The interstitial nuclei of the human anterior hypothalamus: an investigation of variation with sex, sexual orientation, and HIV status. *Horm. Behav.* 40, 86–92.
- Cahill, L., 2006. Why sex matters for neuroscience. *Nat. Rev. Neurosci.* 7, 477–484.
- Cahill, L., Haier, R.J., White, N.S., Fallon, J., Kilpatrick, L., Lawrence, C., Potkin, S.G., Alkire, M.T., 2001. Sex-related difference in amygdala activity during emotionally influenced memory storage. *Neurobiol. Learn. Mem.* 75, 1–9.
- Cahill, L., Uncapher, M., Kilpatrick, L., Alkire, M.T., Turner, J., 2004. Sex-related hemispheric lateralization of amygdala function in emotionally influenced memory: an fMRI investigation. *Learn. Mem.* 11, 261–266.
- Caldwell, L.C., Schweinsburg, A.D., Nagel, B.J., Barlett, V.C., Brown, S.A., Tapert, S.F., 2005. Gender and adolescent alcohol use disorders on BOLD (blood oxygen level dependent) response to spatial working memory. *Alcohol Alcohol.* 40, 194–200.
- Canli, T., Desmond, J.E., Zhao, Z., Gabrieli, J.D., 2002. Sex differences in the neural basis of emotional memories. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10789–10794.
- Carothers, B.J., Reis, H.T., 2013. Men and women are from earth: examining the latent structure of gender. *J. Pers. Soc. Psychol.* 104, 385.
- Carp, J., 2012. The secret lives of experiments: Methods reporting in the fMRI literature. *Neuroimage* 63, 289–300.
- Carrillo, B., Gómez-Gil, E., Rametti, G., Junque, C., Gomez, Á., Karadí, K., Segovia, S., Guillamon, A., 2010. Cortical activation during mental rotation in male-to-female and female-to-male transsexuals under hormonal treatment. *Psychoneuroendocrinology* 35, 1213–1222.
- Caviness Jr, V.S., Kennedy, D.N., Richelme, C., Rademacher, J., Filipek, P.A., 1996. The human brain age 7–11 years: a volumetric analysis based on magnetic resonance images. *Cereb. Cortex* 6, 726–736.
- Ceyhan, M., Adapinar, B., Aksaray, G., Ozdemir, F., Colak, E., 2008. Absence and size of massa intermedia in patients with schizophrenia and bipolar disorder. *Acta Neuropsychiatr.* 20, 193–198.
- Chaplin, T.M., Aldao, A., 2013. Gender differences in emotion expression in children: a meta-analytic review. *Psychol. Bull.* 139, 735.
- Chekroud, A.M., Ward, E.J., Rosenberg, M.D., Holmes, A.V., 2016. Patterns in the human brain mosaic discriminate males from females. *Proc. Natl. Acad. Sci. U. S. A.* 113, E1968.
- Chen, X., Sachdev, P.S., Wen, W., Anstey, K.J., 2007. Sex differences in regional gray matter in healthy individuals aged 44–48 years: a voxel-based morphometric study. *Neuroimage* 36, 691–699.
- Choi, M., Kim, J., Yeon, H., Choi, J., Park, J., Jun, J., Lee, B., Kim, H., Tack, G., Chung, S., 2011. Effects of gender and age on anterior commissure volume. *Neurosci. Lett.* 500, 92–94.
- Christova, P.S., Lewis, S.M., Tagaris, G.A., Ugurbil, K., Georgopoulos, A.P., 2008. A voxel-by-voxel parametric fMRI study of motor mental rotation: hemispheric specialization and gender differences in neural processing efficiency. *Exp. Brain Res.* 189, 79–90.
- Chura, L.R., Lombardo, M.V., Ashwin, E., Auyeung, B., Chakrabarti, B., Bullmore, E.T., Baron-Cohen, S., 2010. Organizational effects of fetal testosterone on human corpus callosum size and asymmetry. *Psychoneuroendocrinology* 35, 122–132.
- Clayton, J.A., 2016. Studying both sexes: a guiding principle for biomedicine. *FASEB J.* 30, 519–524.
- Clements, A.M., Rimrodt, S.L., Abel, J.R., Blankner, J.G., Mostofsky, S.H., Pekar, J.J., Denckla, M.B., Cutting, L.E., 2006. Sex differences in cerebral laterality of language and visuospatial processing. *Brain Lang.* 98, 150–158.
- Clements-Stephens, A., Rimrodt, S.L., Cutting, L.E., 2009. Developmental sex differences in basic visuospatial processing: differences in strategy use? *Neurosci. Lett.* 449, 155–160.
- Cohen, J., 1992. A power primer. *Psych. Bull.* 112, 155–159.
- Colaer, M.L., Reimers, S., Manning, J.T., 2007. Visuospatial performance on an internet line judgment task and potential hormonal markers: sex, sexual orientation, and 2D: 4D. *Arch. Sex. Behav.* 36, 177–192.
- Cosgrove, K.P., Mazure, C.M., Staley, J.K., 2007. Evolving knowledge of sex differences in brain structure, function, and chemistry. *Biol. Psychiatry* 62, 847–855.
- Coupé, P., Catheline, G., Lanuza, E., Manjón, J.V., 2017. Towards a unified analysis of brain maturation and aging across the entire lifespan: a MRI analysis. *Hum. Brain Mapp.* 38, 5501–5518.
- Cox, S.R., McKenzie, T.I., Arribasal, B.S., Royle, N.A., MacPherson, S.E., MacLullich, A.M. J., Bastin, M.E., Wardlaw, J.M., Deary, I.J., Ferguson, K.J., 2016. Volumetric and correlational implications of brain parcellation method selection: a 3-way comparison in the frontal lobes. *J. Comput. Assist. Tomogr.* 40, 53–60.
- Crespo-Facorro, B., Roiz-Santínez, R., Pérez-Iglesias, R., Mata, I., Rodríguez-Sánchez, J. M., Tordesillas-Gutiérrez, D., de, I.F., Tabarés-Seisdedos, R., Sánchez, E., Andreasen, N., Magnotta, V., Vázquez-Barquero, J.L., 2011. Sex-specific variation of MRI-based cortical morphometry in adult healthy volunteers: The effect on cognitive functioning. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 616–623.

- Creze, M., Verschueren, L., Besson, P., Sauvage, C., Leclerc, X., Jissendi-Tchofo, P., 2014. Age- and gender-related regional variations of human brain cortical thickness, complexity, and gradient in the third decade. *Hum. Brain Mapp.* 35, 2817–2835.
- Damle, N.R., Ikuta, T., John, M., Peters, B.D., DeRosse, P., Malhotra, A.K., Szczek, P.R., 2017. Relationship among interthalamic adhesion size, thalamic anatomy and neuropsychological functions in healthy volunteers. *Brain Struct. Funct.* 222, 2183–2192.
- David, S.P., Naudet, F., Laude, J., Radua, J., Fusar-Poli, P., Chu, I., Stefanick, M.L., Ioannidis, J.P., 2018. Potential reporting bias in neuroimaging studies of sex differences. *Sci. Rep.* 8, 6082.
- Davis, E.P., Pfaff, D., 2014. Sexually dimorphic responses to early adversity: implications for affective problems and autism spectrum disorder. *Psychoneuroendocrinology* 49, 11–25.
- de Jong, L.W., Vidal, J.S., Forsberg, L.E., Zijdenbos, A.P., Haight, T., Alzheimer's Disease Neuroimaging Initiative, Sigurdsson, S., Gudnason, V., van Buchem, M.A., Launer, L.J., 2017. Allometric scaling of brain regions to intra-cranial volume: an epidemiological MRI study. *Hum. Brain Mapp.* 38, 151–164.
- de la Grandmaison, G.L., Clairand, I., Durigon, M., 2001. Organ weight in 684 adult autopsies: new tables for a caucasian population. *Forensic Sci. Int.* 119, 149–154.
- De Lacoste, M.C., Holloway, R.L., Woodward, D.J., 1986. Sex differences in the fetal human corpus callosum. *Hum. Neurobiol.* 5, 93–96.
- De Lacy, N., Calhoun, V.D., Kutz, J.N., McCauley, E., 2019. Multilevel mapping of sexual dimorphism in intrinsic functional brain networks. *Front. Neurosci.* 13, 332.
- de Souza Crippa, J.A., Zuardi, A.W., Busatto, G.F., Sanches, R.F., Santos, A.C., Araujo, D., Amaro, E., Hallak, J.E., Ng, V., McGuire, P.K., 2006. Cavum septum pellucidum and adhesio interthalamicia in schizophrenia: an MRI study. *Eur. Psychiatry* 21, 291–299.
- De Vries, G.J., 2004. Minireview: sex differences in adult and developing brains: compensation, compensation, compensation. *Endocrinology* 145, 1063–1068.
- De Vries, G.J., Södersten, P., 2009. Sex differences in the brain: the relation between structure and function. *Horm. Behav.* 55, 589–596.
- Dean, D.C., Planalp, E.M., Wooten, W., Schmidt, C.K., Kecskemeti, S.R., Frye, C., Schmidt, N.L., Goldsmith, H.H., Alexander, A.L., Davidson, R.J., 2018. Investigation of brain structure in the 1-month infant. *Brain Struct. Funct.* 223, 1953–1970.
- DeLaCoste-Utamis, C., Holloway, R.L., 1982. Sexual dimorphism in the human corpus callosum. *Science* 216, 1431–1432.
- Demeter, S., Ringo, J.L., Doty, R.W., 1988. Morphometric analysis of the human corpus callosum and anterior commissure. *Hum. Neurobiol.* 6, 219–226.
- Dennis, E.L., Jahanshad, N., McMahon, K.L., de Zubicaray, G.I., Martin, N.G., Hickie, I.B., Toga, A.W., Wright, M.J., Thompson, P.M., 2013. Development of brain structural connectivity between ages 12 and 30: a 4-tesla diffusion imaging study in 439 adolescents and adults. *Neuroimage* 64, 671–684.
- Deslauriers, M., 2009. Sexual difference in Aristotle's politics and his biology. *Class. World* 215–231.
- Draganski, B., Gaser, C., Kempermann, G., Kuhn, H.G., Winkler, J., Buchel, C., May, A., 2006. Temporal and spatial dynamics of brain structure changes during extensive learning. *J. Neurosci.* 26, 6314–6317.
- Driesen, N.R., Raz, N., 1995. The influence of sex, age, and handedness on corpus callosum morphology: a meta-analysis. *Psychobiology* 23, 240–247.
- Drobyshevsky, A., Baumann, S.B., Schneider, W., 2006. A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function. *Neuroimage* 31, 732–744.
- Duarte-Carvalhalino, J., Jahanshad, N., Lenglet, C., McMahon, K.L., de Zubicaray, G.I., Martin, N.G., Wright, M.J., Thompson, P.M., Sapiro, G., 2012. Hierarchical topological network analysis of anatomical human brain connectivity and differences related to sex and kinship. *Neuroimage* 59, 3784–3804.
- Durston, S., Hulshoff Pol, H.E., Casey, B.J., Giedd, J.N., Buitelaar, J.K., van Engeland, H., 2001. Anatomical MRI of the developing human brain: what have we learned? *J. Am. Acad. Child Adolesc. Psychiatry* 40, 1012–1020.
- Eliot, L., 2011. The trouble with sex differences. *Neuron* 72, 895–898.
- Elliott, M.L., Knott, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., Sison, M.L., Moffitt, T.E., Caspi, A., Hariri, A.R., 2020. What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychol. Sci.* 31 (7), 792–806.
- Else-Quest, N.M., Higgins, A., Allison, C., Morton, L.C., 2012. Gender differences in self-conscious emotional experience: a meta-analysis. *Psychol. Bull.* 138, 947.
- Erbagci, H., Yildirim, H., Herken, H., Gümrüksurun, E., 2002. A magnetic resonance imaging study of the adhesio interthalamicia in schizophrenia. *Schizophr. Res.* 55, 89–92.
- Erickson, K.I., Voss, M.W., Prakash, R.S., Basak, C., Szabo, A., Chaddock, L., Kim, J.S., Heo, S., Alves, H., White, S.M., Wojcicki, T.R., Mailey, E., Vieira, V.J., Martin, S.A., Pence, B.D., Woods, J.A., McAuley, E., Kramer, A.F., 2011. Exercise training increases size of hippocampus and improves memory. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3017–3022.
- Etchell, A., Adhikari, A., Weinberg, L.S., Choo, A.L., Garnett, E.O., Chow, H.M., Chang, S.E., 2018. A systematic literature review of sex differences in childhood language and brain development. *Neuropsychologia* 114, 19–31.
- Feis, D., Brodersen, K.H., von Cramon, D.Y., Luders, E., Tittgemeyer, M., 2013. Decoding gender dimorphism of the human brain using multimodal anatomical and diffusion MRI data. *Neuroimage* 70, 250–257.
- Ferretti, M.T., Iulita, M.F., Cavedo, E., Chiesa, P.A., Schumacher Dimech, A., Santuccione Chadha, A., Baracchi, F., Girouard, H., Misoch, S., Giacobini, E., Depypere, H., Hampel, H., Women's Brain Project and the Alzheimer Precision Medicine Initiative, 2018. Sex differences in Alzheimer disease - the gateway to precision medicine. *Nat. Rev. Neurol.* 14, 457–469.
- Filipek, P.A., Richelme, C., Kennedy, D.N., Caviness Jr, V.S., 1994. The young adult human brain: an MRI-based morphometric analysis. *Cereb. Cortex* 4, 344–360.
- Filippi, M., Valsasina, P., Misci, P., Falini, A., Comi, G., Rocca, M.A., 2013. The organization of intrinsic brain activity differs between genders: a resting-state fMRI study in a large cohort of young healthy subjects. *Hum. Brain Mapp.* 34, 1330–1343.
- Filkowski, M.M., Olsen, R.M., Duda, B., Wanger, T.J., Sabatinelli, D., 2017. Sex differences in emotional perception: meta analysis of divergent activation. *Neuroimage* 147, 925–933.
- Fine, C., 2010. From scanner to sound bite: issues in interpreting and reporting sex differences in the brain. *Curr. Dir. Psychol. Sci.* 19, 280–283.
- Fjell, A.M., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., Dale, A.M., Walhovd, K.B., 2009. Minute effects of sex on the aging brain: a multisample magnetic resonance imaging study of healthy aging and Alzheimer's disease. *J. Neurosci.* 29, 8774–8783.
- Forstmeier, W., 2011. Women have relatively larger brains than men: a comment on the misuse of general linear models in the study of sexual dimorphism. *Anat. Rec.* 294, 1856–1863.
- Fotuhi, M., Do, D., Jack, C., 2012. Modifiable factors that alter the size of the hippocampus with ageing. *Nat. Rev. Neurol.* 8, 189–202.
- Free, S.L., Bergin, P.S., Fish, D.R., Cook, M.J., Shorvon, S.D., Stevens, J.M., 1995. Methods for normalization of hippocampal volumes measured with MR. *AJR Am. J. Neuroradiol.* 16, 637–643.
- Frick, A., Möhring, W., Newcombe, N.S., 2014. Development of mental transformation abilities. *Trends Cogn. Sci.* 18, 536–542.
- Frost, J.A., Binder, J.R., Springer, J.A., Hammeken, T.A., Bellgowan, P.S.F., Rao, S.M., Cox, R.W., 1999. Language processing is strongly left lateralized in both sexes: evidence from functional MRI. *Brain* 122, 199–208.
- Gaillard, W.D., Sachs, B.C., Whitnah, J.R., Ahmad, Z., Balsamo, L.M., Petrella, J.R., Branicki, S.H., McKinney, C.M., Hunter, K., Xu, B., Grandin, C.B., 2003. Developmental aspects of language processing: fMRI of verbal fluency in children and adults. *Hum. Brain Mapp.* 18, 176–185.
- Garcia-Falgueras, A., Swaab, D.F., 2008. A sex difference in the hypothalamic uncinate nucleus: relationship to gender identity. *Brain* 131, 3132–3146.
- Garcia-Garcia, I., Kube, J., Gaebler, M., Horstmann, A., Villringer, A., Neumann, J., 2016. Neural processing of negative emotional stimuli and the influence of age, sex and task-related characteristics. *Neurosci. Biobehav. Rev.* 68, 773–793.
- Garn, C.L., Allen, M.D., Larsen, J.D., 2009. An fMRI study of sex differences in brain activation during object naming. *Cortex* 45, 610–618.
- Gauthier, C.T., Duyme, M., Zanca, M., Capron, C., 2009. Sex and performance level effects on brain activation during a verbal fluency task: a functional magnetic resonance imaging study. *Cortex* 45, 164–176.
- Geary, D.C., 2019. Evolutionary perspective on sex differences in the expression of neurological diseases. *Prog. Neurobiol.* 176, 33–53.
- Geng, X., Li, G., Lu, Z., Gao, W., Wang, L., Shen, D., Zhu, H., Gilmore, J.H., 2017. Structural and maturational covariance in early childhood brain development. *Cereb. Cortex* 27, 1795–1807.
- Gennatas, E.D., Avants, B.B., Wolf, D.H., Satterthwaite, T.D., Ruparel, K., Ciric, R., Hakonarson, H., Gur, R.E., Gur, R.C., 2017. Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. *J. Neurosci.* 37, 5065–5073.
- Geschwind, N., Behan, P., 1982. Left-handedness: association with immune disease, migraine, and developmental learning disorder. *Proc. Natl. Acad. Sci. U. S. A.* 79, 5097–5100.
- Geschwind, N., Galaburda, A.M., 1985a. Cerebral lateralization: biological mechanisms, associations, and pathology: I. A hypothesis and a program for research. *Arch. Neurol.* 42, 428–459.
- Geschwind, N., Galaburda, A.M., 1985b. Cerebral lateralization: biological mechanisms, associations, and pathology: II. A hypothesis and a program for research. *Arch. Neurol.* 42, 521–552.
- Geschwind, N., Galaburda, A.M., 1985c. Cerebral lateralization: biological mechanisms, associations, and pathology: III. A hypothesis and a program for research. *Arch. Neurol.* 42, 634–654.
- Giedd, J.N., Vaituzis, A.C., Hamburger, S.D., Lange, N., Rajapakse, J.C., Kayser, D., Vauss, Y.C., Rapoport, J.L., 1996. Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development: ages 4–18 years. *J. Comp. Neurol.* 366, 223–230.
- Giedd, J.N., Castellanos, F.X., Rajapakse, J.C., Vaituzis, A.C., Rapoport, J.L., 1997. Sexual dimorphism of the developing human brain. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 21, 1185–1201.
- Giedd, J.N., Raznahan, A., Mills, K.L., Lenroot, R.K., 2012. Review: magnetic resonance imaging of male/female differences in human adolescent brain anatomy. *Biol. Sex Diff.* 3, 19–27.
- Gizewski, E.R., Krause, E., Wanke, I., Forsting, M., Senf, W., 2006. Gender-specific cerebral activation during cognitive tasks using functional MRI: comparison of women in mid-luteal phase and men. *Neuroradiology* 48, 14–20.
- Goldman, B., 2017. Two Minds: The Cognitive Differences Between Men and Women. Stanford Medicine, p. 12. <http://stanmed.stanford.edu/2017spring/how-mens-and-womens-brains-are-different.html>.
- Goldstein, J.M., Seidman, L.J., Horton, N.J., Makris, N., Kennedy, D.N., Caviness, V.S.J., Faraone, S.V., Tsuang, M.T., 2001. Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging. *Cereb. Cortex* 11, 490–497.
- Goldstein, J.M., Jerram, M., Poldrack, R., Ahern, T., Kennedy, D.N., Seidman, L.J., Makris, N., 2005. Hormonal cycle modulates arousal circuitry in women using functional magnetic resonance imaging. *J. Neurosci.* 25, 9309–9316.
- Gong, G., Rosa-Neto, P., Carbonell, F., Chen, Z.J., He, Y., Evans, A.C., 2009. Age- and gender-related differences in the cortical anatomical network. *J. Neurosci.* 29, 15684–15693.

- Gong, G., He, Y., Evans, A.C., 2011. Brain connectivity: gender makes a difference. *Neuroscientist* 17, 575–591.
- Good, C.D., Johnsrude, I., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *Neuroimage* 14, 685–700.
- Goodro, M., Sameti, M., Patenaude, B., Fein, G., 2012. Age effect on subcortical structures in healthy adults. *Psychiatry Res. Neuroimaging* 203, 38–45.
- Gorski, R.A., Gordon, J.H., Shryne, J.E., Southam, A.M., 1978. Evidence for a morphological sex difference within the medial preoptic area of the rat brain. *Brain Res.* 148, 333–346.
- Govindarajan, K.A., Freeman, L., Cai, C., Rahbar, M.H., Narayana, P.A., 2014. Effect of intrinsic and extrinsic factors on global and regional cortical thickness. *PLoS One* 9, e96429.
- Grabowska, A., 2017. Sex on the brain: are gender-dependent structural and functional differences associated with behavior? *J. Neurosci. Res.* 95, 200–212.
- Grant, B.F., Dawson, D.A., Stinson, F.S., Chou, S.P., Dufour, M.C., Pickering, R.P., 2004. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United states, 1991–1992 and 2001–2002. *Drug Alcohol Depend.* 74, 223–234.
- Grimm, O., Pohlack, S., Cacciaglia, R., Winkelmann, T., Plichta, M.M., Demirakca, T., Flor, H., 2015. Amygdala and hippocampal volume: a comparison between manual segmentation, freesurfer and VBM. *J. Neurosci. Methods* 253, 254–261.
- Grön, G., Wunderlich, A.P., Spitzer, M., Tomczak, R., Riepe, M.W., 2000. Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nat. Neurosci.* 3, 404–408.
- Guadalupe, T., Zwiers, M.P., Wittfeld, K., Teumer, A., Vasquez, A.A., Hoogman, M., Hagoort, P., Fernandez, G., Buitelaar, J., van Bokhoven, H., 2015. Asymmetry within and around the human planum temporale is sexually dimorphic and influenced by genes involved in steroid hormone receptor activity. *Cortex* 62, 41–55.
- Gur, R.E., Gur, R.C., 2016. Sex differences in brain and behavior in adolescence: findings from the Philadelphia neurodevelopmental cohort. *Neurosci. Biobehav. Rev.* 70, 159–170.
- Gur, R.C., Alsop, D., Glahn, D., Petty, R., Swanson, C.L., Maldjian, J.A., Turetsky, B.I., Detre, J.A., Gee, J., Gur, R.E., 2000. An fMRI study of sex differences in regional activation to a verbal and a spatial task. *Brain Lang.* 74, 157–170.
- Gur, R.C., Gunning-Dixon, F., Bilker, W.B., Gur, R.E., 2002. Sex differences in temporolimbic and frontal brain volumes of healthy adults. *Cereb. Cortex* 12, 998–1003.
- Gur, R.C., Bockow, T., Gur, R.E., 2010. Gender differences in the functional organization of the brain. *Principles of Gender-specific Medicine Gender Differences in the Functional Organization of the Brain*. Elsevier, pp. 75–86.
- Gurian, M., Stevens, K., 2004. With boys and girls in mind. *Educ. Leadership* 62, 21–26.
- Halari, R., Sharma, T., Hines, M., Andrew, C., Simmons, A., Kumari, V., 2006. Comparable fMRI activity with differential behavioural performance on mental rotation and overt verbal fluency tasks in healthy men and women. *Exp. Brain Res.* 169, 1–14.
- Hall, J.A., 1978. Gender effects in decoding nonverbal cues. *Psychol. Bull.* 85, 845–857.
- Halpern, D.F., 2012. *Sex Differences in Cognitive Abilities*, 4th ed. Psychology Press, New York.
- Hamann, S., 2005. Sex differences in the responses of the human amygdala. *Neuroscientist* 11, 288–293.
- Hammerslag, L.R., Gulley, J.M., 2016. Sex differences in behavior and neural development and their role in adolescent vulnerability to substance use. *Behav. Brain Res.* 298, 15–26.
- Hänggi, J., Buchmann, A., Mondadori, C.R.A., Henke, K., Jäncke, L., Hock, C., 2010. Sexual dimorphism in the parietal substrate associated with visuospatial cognition independent of general intelligence. *J. Cogn. Neurosci.* 22, 139–155.
- Hänggi, J., Fovényi, L., Liem, F., Meyer, M., Jäncke, L., 2014. The hypothesis of neuronal interconnectivity as a function of brain size—a general organization principle of the human connectome. *Front. Hum. Neurosci.* 8, 915.
- Harrington, G.S., Farias, S.T., 2008. Sex differences in language processing: functional MRI methodological considerations. *J. Magn. Reson. Imaging* 27, 1221–1228.
- Hattemer, K., Plate, A., Heverhagen, J.T., Haag, A., Keil, B., Klein, K.M., Hermsen, A., Oertel, W.H., Hamer, H.M., Rosenow, F., Knake, S., 2011. Determination of hemispheric dominance with mental rotation using functional transcranial Doppler sonography and fMRI. *J. Neuroimaging* 21, 16–23.
- Hausmann, M., 2017. Why sex hormones matter for neuroscience: a very short review on sex, sex hormones, and functional brain asymmetries. *J. Neurosci. Res.* 95, 40–49.
- Haut, K.M., Barch, D.M., 2006. Sex influences on material-sensitive functional lateralization in working and episodic memory: men and women are not all that different. *Neuroimage* 32, 411–422.
- Hawes, Z., Sokolowski, H.M., Ononye, C.B., Ansari, D., 2019. Neural underpinnings of numerical and spatial cognition: an fMRI meta-analysis of brain regions associated with symbolic number, arithmetic, and mental rotation. *Neurosci. Biobehav. Rev.* 103, 316–336.
- Herron, T.J., Kang, X., Woods, D.L., 2012. Automated measurement of the human corpus callosum using MRI. *Front. Neuroinform.* 6, 25.
- Herting, M.M., Johnson, C., Mills, K.L., Vijayakumar, N., Dennison, M., Liu, C., Goddins, A., Dahl, R.E., Sowell, E.R., Whittle, S., Allen, N.B., Tamnes, C.K., 2018. Development of subcortical volumes across adolescence in males and females: a multisample study of longitudinal changes. *Neuroimage* 172, 194–205.
- Highley, J.R., Esiri, M.M., McDonald, B., Roberts, H.C., Walker, M.A., Crow, T.J., 1999. The size and fiber composition of the anterior commissure with respect to gender and schizophrenia. *Biol. Psychiatry* 45, 1120–1127.
- Hines, M., 2011. Gender development and the human brain. *Ann. Rev. Neurosci.* 34, 69–88.
- Hirnstein, M., Hugdahl, K., Hausmann, M., 2019. Cognitive sex differences and hemispheric asymmetry: a critical review of 40 years of research. *Laterality* 24, 204–252.
- Hiscock, M., Inch, R., Jacek, C., Hiscock-Kalil, C., Kalil, K.M., 1994. Is there a sex difference in human laterality? I. An exhaustive survey of auditory laterality studies from six neuropsychology journals. *J. Clin. Exp. Neuropsychol.* 16, 423–435.
- Hiscock, M., Israeli, M., Inch, R., Jacek, C., Hiscock-Kalil, C., 1995. Is there a sex difference in human laterality? II. An exhaustive survey of visual laterality studies from six neuropsychology journals. *J. Clin. Exp. Neuropsychol.* 17, 590–610.
- Hiscock, M., Inch, R., Hawryluk, J., Lyon, P.J., Perachio, N., 1999. Is there a sex difference in human laterality? III. An exhaustive survey of tactile laterality studies from six neuropsychology journals. *J. Clin. Exp. Neuropsychol.* 21, 17–28.
- Hiscock, M., Perachio, N., Inch, R., 2001. Is there a sex difference in human laterality? IV. An exhaustive survey of dual-task interference studies from six neuropsychology journals. *J. Clin. Exp. Neuropsychol.* 23, 137–148.
- Holloway, R.L., Anderson, P.J., Defendini, R., Harper, C., 1993. Sexual dimorphism of the human corpus callosum from three independent samples: relative size of the corpus callosum. *Am. J. Phys. Anthropol.* 92, 481–498.
- Hong, Y.W., Yoo, Y., Han, J., Wager, T.D., Woo, C.W., 2019. False-positive neuroimaging: undisclosed flexibility in testing spatial hypotheses allows presenting anything as a replicated finding. *Neuroimage* 195, 384–395.
- Hoppe, C., Fliessbach, K., Stausberg, S., Stojanovic, J., Trautner, P., Elger, C.E., Weber, B., 2012. A key role for experimental task performance: effects of math talent, gender and performance on the neural correlates of mental rotation. *Brain Cogn.* 78, 14–27.
- Hudson, J.I., Hiripi, E., Pope, H.G., Kessler, R.C., 2007. The prevalence and correlates of eating disorders in the national comorbidity survey replication. *Biol. Psychiatry* 61, 348–358.
- Hufner, K., Binetti, C., Hamilton, D.A., Stephan, T., Flanagan, V.L., Linn, J., Labudda, K., Markowitz, H., Glasauer, S., Jahn, K., Strupp, M., Brandt, T., 2011. Structural and functional plasticity of the hippocampal formation in professional dancers and slackliners. *Hippocampus* 21, 855–865.
- Hugdahl, K., Thomsen, T., Ersland, L., 2006. Sex differences in visuo-spatial processing: an fMRI study of mental rotation. *Neuropsychologia* 44, 1575–1583.
- Hyde, J.S., 2005. The gender similarities hypothesis. *Am. Psychol.* 60, 581–592.
- Hyde, J.S., 2014. Gender similarities and differences. *Annu. Rev. Psychol.* 65, 373–398.
- Hyde, J.S., 2016. Sex and cognition: gender and cognitive functions. *Curr. Opin. Neurobiol.* 38, 53–56.
- Hyde, J.S., Linn, M.C., 1988. Gender differences in verbal ability: a meta-analysis. *Psychol. Bull.* 104, 53–69.
- Ihnen, S.K.Z., Church, J.A., Petersen, S.E., Schlaggar, B.L., 2009. Lack of generalizability of sex differences in the fMRI BOLD activity associated with language processing in adults. *Neuroimage* 45, 1020–1032.
- Im, K., Lee, J., Lee, J., Shin, Y., Kim, I.Y., Kwon, J.S., Kim, S.I., 2006. Gender difference analysis of cortical thickness in healthy young adults with surface-based methods. *Neuroimage* 31, 31–38.
- Im, K., Lee, J., Lyttelton, O., Kim, S.H., Evans, A.C., Kim, S.I., 2008. Brain size and cortical structure in the adult human brain. *Cereb. Cortex* 18, 2181–2191.
- Inano, S., Takao, H., Hayashi, N., Yoshioka, N., Mori, H., Kunimatsu, A., Abe, O., Ohtomo, K., 2013. Effects of age and gender on neuroanatomical volumes. *J. Magn. Reson. Imaging* 37, 1072–1076.
- Ingalahikar, M., Smith, A., Parker, D., Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Hakanson, H., Gur, R.E., Gur, R.C., Verma, R., 2014. Sex differences in the structural connectome of the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 111, 823–828.
- Ioannidis, J.P.A., 2011. Excess significance bias in the literature on brain volume abnormalities. *Arch. Gen. Psychiatry* 68, 773–780.
- Ioannidis, J.P.A., Munafò, M.R., Fusar-Poli, P., Nosek, B.A., David, S.P., 2014. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends Cogn. Sci.* 18, 235–241.
- Jack Jr, C.R., Twomey, C.K., Zinsmeister, A.R., Sharbrough, F.W., Petersen, R.C., Cascino, G.D., 1989. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. *Radiology* 172, 549–554.
- Jäncke, L., Staiger, J.F., Schlaug, G., Huang, Y., Steinmetz, H., 1997. The relationship between corpus callosum size and forebrain volume. *Cereb. Cortex* 7, 48–56.
- Jäncke, L., Mérillat, S., Liem, F., Hänggi, J., 2015. Brain size, sex, and the aging brain. *Hum. Brain Mapp.* 36, 150–169.
- Janz, K.F., Dawson, J.D., Mahoney, L.T., 2000. Predicting heart growth during puberty: the muscatine study. *Pediatrics* 105, E63.
- Jazin, E., Cahill, L., 2010. Sex differences in molecular neuroscience: from fruit flies to humans. *Nat. Rev. Neurosci.* 11, 9–17.
- Jensen, F.E., Nutt, A.E., 2015. Teen girls have different brains: gender, neuroscience and the truth about adolescence. *Salon*, 3-Jan-2015. https://www.salon.com/2015/01/03/teen_girls_have_different_brains_gender_neuroscience_and_the_truth_about_adolescence/.
- Joel, D., 2011. Male or female? Brains are intersex. *Front. Integr. Neurosci.* 5, 1–5.
- Joel, D., 2012. Genetic-gonadal-genitals sex (3G-sex) and the misconception of brain and gender, or, why 3G-males and 3G-females have intersex brain and intersex gender. *Biol. Sex Diff.* 3, 27.
- Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., Shefi, N., Pool, J., Urchs, S., Margulies, D.S., Liem, F., Hänggi, J., Jancke, L., Assaf, Y., 2015. Sex beyond the genitalia: the human brain mosaic. *Proc. Natl. Acad. Sci. U. S. A.* 112, 15468–15473.

- Joel, D., Persico, A., Salhov, M., Berman, Z., Oligschläger, S., Meilijson, I., Averbuch, A., 2018. Analysis of human brain structure reveals that the brain "Types" typical of males are also typical of females, and vice versa. *Front. Hum. Neurosci.* 12, 399.
- Jones, C.M., Braithwaite, V.A., Healy, S.D., 2003. The evolution of sex differences in spatial ability. *Behav. Neurosci.* 117, 403.
- Jordan, K., Wißnerberg, T., Heinze, H.J., Peters, M., Jäncke, L., 2002. Women and men exhibit different cortical activation patterns during mental rotation tasks. *Neuropsychologia* 40, 2397–2408.
- Jordan-Young, R., Rumiantseva, R.I., 2012. Hardwired for sexism? Approaches to s/g in neuroscience. *Neuroethics* 5, 305–315.
- Kaczurkin, A.N., Raznahan, A., Satterthwaite, T.D., 2019. Sex differences in the developing brain: insights from multimodal neuroimaging. *Neuropsychopharmacology* 44, 71–85.
- Kaiser, A., Kuenzli, E., Zappatorta, D., Nitsch, C., 2007. On females' lateral and males' bilateral activation during language production: a fMRI study. *Int. J. Psychophysiol.* 63, 192–198.
- Kaiser, S., Walther, S., Nennig, E., Kronmuller, K., Mundt, C., Weisbrod, M., Stippich, C., Vogeley, K., 2008. Gender-specific strategy use and neural correlates in a spatial perspective-taking task. *Neuropsychologia* 46, 2524–2531.
- Kaiser, A., Haller, S., Schmitz, S., Nitsch, C., 2009. On sex/gender related similarities and differences in fMRI language research. *Brain Res. Rev.* 61, 49–59.
- Kana, R.K., Murdaugh, D.L., Wolfe, K.R., Kumar, S.L., 2012. Brain responses mediating idiom comprehension: gender and hemispheric differences. *Brain Res.* 1467, 18–26.
- Kansaku, K., Yamaura, A., Kitazawa, S., 2000. Sex differences in lateralization revealed in the posterior language areas. *Cereb. Cortex* 10, 866–872.
- Keller, K., Menon, V., 2009. Gender differences in the functional and structural neuroanatomy of mathematical cognition. *Neuroimage* 47, 342–352.
- Kennedy, K.M., Erickson, K.I., Rodriguez, K.M., Voss, M.W., Colcombe, S.J., Kramer, A.F., Acker, J.D., Raz, N., 2009. Age-related differences in regional brain volumes: a comparison of optimized voxel-based morphometry to manual volumetry. *Neurobiol. Aging* 30, 1657–1676.
- Király, A., Szabó, N., Tóth, E., Csete, G., Faragó, P., Kocsis, K., Must, A., Vécsei, L., Kincses, Z.T., 2016. Male brain ages faster: the age and gender dependence of subcortical volumes. *Brain Imaging Behav.* 10, 901–910.
- Knickmeyer, R.C., Wang, J., Zhu, H., Geng, X., Woolson, S., Hamer, R.M., Konneker, T., Styner, M., Gilmore, J.H., 2014. Impact of sex and gonadal steroids on neonatal brain structure. *Cereb. Cortex* 24, 2721–2731.
- Knickmeyer, R.C., Xia, K., Lu, Z., Ahn, M., Jha, S.C., Zou, F., Zhu, H., Styner, M., Gilmore, J.H., 2017. Impact of demographic and obstetric factors on infant brain volumes: a population neuroscience study. *Cereb. Cortex* 27, 5616–5625.
- Kocak, M., Ulmer, J.L., Biswal, B.B., Aralsmalk, A., Daniels, D.L., Mark, L.P., 2005. The influence of gender on auditory and language cortical activation patterns: preliminary data. *AJNR Am. J. Neuroradiol.* 26, 2248–2255.
- Kochunov, P., et al., 2015. Heritability of fractional anisotropy in human white matter: a comparison of human connectome project and ENIGMA-DTI data. *Neuroimage* 111, 300–311.
- Konrad, C., Engelien, A., Schoning, S., Zwitserlood, P., Jansen, A., Pletziger, E., Beizai, P., Kersting, A., Ohrmann, P., Luders, E., Greb, R.R., Heindel, W., Arolt, V., Kugel, H., 2008. The functional anatomy of semantic retrieval is influenced by gender, menstrual cycle, and sex hormones. *J. Neural Transm.* 115, 1327–1337.
- Koscik, T., O'Leary, D., Moser, D.J., Andreassen, N.C., Nopoulos, P., 2009. Sex differences in parietal lobe morphology: relationship to mental rotation performance. *Brain Cogn.* 69, 451–459.
- Kruggel, F., 2006. MRI-based volumetry of head compartments: normative values of healthy adults. *Neuroimage* 30, 1–11.
- Kucian, K., Loenneker, T., Dietrich, T., Martin, E., Von Aster, M., 2005. Gender differences in brain activation patterns during mental rotation and number related cognitive tasks. *Psychol. Sci.* 47, 112–131.
- Kucian, K., Von Aster, M., Loenneker, T., Dietrich, T., Mast, F.W., Martin, E., 2007. Brain activation during mental rotation in school children and adults. *J. Neural Transm.* 114, 675–686.
- Lammin, C., Decety, J., Singer, T., 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502.
- Larsen, J.D., Allen, M.D., Bigler, E.D., Goodrich-Hunsaker, N.J., Hopkins, R.O., 2010. Different patterns of cerebral activation in genuine and malingered cognitive effort during performance on the word memory test. *Brain Inj.* 24, 89–99.
- Lasco, M.S., Jordan, T.J., Edgar, M.A., Petito, C.K., Byne, W., 2002. A lack of dimorphism of sex or sexual orientation in the human anterior commissure. *Brain Res.* 936, 95–98.
- Lauer, J.E., Yhang, E., Lourenco, S.F., 2019. The development of gender differences in spatial reasoning: a meta-analytic review. *Psychol. Bull.* 145, 537–565.
- Lemaître, H., Crivello, F., Grassiot, B., Alpérovitch, A., Tzourio, C., Mazoyer, B., 2005. Age- and sex-related effects on the neuroanatomy of healthy elderly. *Neuroimage* 26, 900–911.
- Lenroot, R.K., Gogtay, N., Greenstein, D.K., Wells, E.M., Wallace, G.L., Clasen, L.S., Blumenthal, J.D., Lerch, J., Zijdenbos, A.P., Evans, A.C., Thompson, P.M., Giedd, J.N., 2007. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage* 36, 1065–1073.
- Leonard, C.M., Towler, S., Welcome, S., Halderman, L.K., Otto, R., Eckert, M.A., Chiarello, C., 2008. Size matters: cerebral volume influences sex differences in neuroanatomy. *Cereb. Cortex* 18, 2920–2931.
- LeVay, S., 1991. A difference in hypothalamic structure between heterosexual and homosexual men. *Science* 253, 1034.
- Lim, S., Han, C.E., Uhlhaas, P.J., Kaiser, M., 2015. Preferential detachment during human brain development: age- and sex-specific structural connectivity in diffusion tensor imaging (DTI) data. *Cereb. Cortex* 25, 1477–1489.
- Linn, M.C., Petersen, A.C., 1985. Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev.* 56, 1479–1498.
- Lippa, R.A., 2002. Gender, Nature, and Nurture. Lawrence Erlbaum Associates, Mahwah, NJ.
- Liu, H., Stufflebeam, S.M., Sepulcre, J., Hedden, T., Buckner, R.L., 2009. Evidence from intrinsic activity that asymmetry of the human brain is controlled by multiple factors. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20499–20503.
- Liu, D., Johnson, H.J., Long, J.D., Magnotta, V.A., Paulsen, J.S., 2014. The power-proportion method for intracranial volume correction in volumetric imaging analysis. *Front. Neurosci.* 8, 356.
- Liu, S., Seidlitz, J., Blumenthal, J.D., Clasen, L.S., Raznahan, A., 2020. Integrative structural, functional, and transcriptomic analyses of sex-biased brain organization in humans. *Proc. Natl. Acad. Sci. U. S. A.* 117, 18788–18798.
- Loomes, R., Hull, L., Mandy, W.P.L., 2017. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry* 56, 466–474.
- Lotze, M., Domin, M., Gerlach, F.H., Gaser, C., Lueders, E., Schmidt, C.O., Neumann, N., 2019. Novel findings from 2,838 adult brains on sex differences in gray matter brain volume. *Sci. Rep.* 9, 1–7.
- Luders, E., Toga, A.W., 2010. Sex differences in brain anatomy. *Prog. Brain Res.* 186, 3–12.
- Luders, E., Steinmetz, H., Jancke, L., 2002. Brain size and grey matter volume in the healthy human brain. *Neuroreport* 13, 2371–2374.
- Luders, E., Narr, K.L., Thompson, P.M., Rex, D.E., Woods, R.P., DeLuca, H., Jäncke, L., Toga, A.W., 2006a. Gender effects on cortical thickness and the influence of scaling. *Hum. Brain Mapp.* 27, 314–324.
- Luders, E., Narr, K.L., Zaidel, E., Thompson, P.M., Toga, A.W., 2006b. Gender effects on callosal thickness in scaled and unscaled space. *Neuroreport* 17, 1103–1106.
- Luders, E., Gaser, C., Narr, K.L., Toga, A.W., 2009. Why sex matters: brain size independent differences in gray matter distributions between men and women. *J. Neurosci.* 29, 14265–14270.
- Luders, E., Sánchez, F.J., Tosun, D., Shattuck, D.W., Gaser, C., Vilain, E., Toga, A.W., 2012. Increased cortical thickness in male-to-female transsexualism. *J. Behav. Brain Sci.* 2, 357–362.
- Luders, E., Thompson, P.M., Kurth, F., Hong, J., Phillips, O.R., Wang, Y., Gutman, B.A., Chou, Y., Narr, K.L., Toga, A.W., 2013. Global and regional alterations of hippocampal anatomy in long-term meditation practitioners. *Hum. Brain Mapp.* 34, 3369–3375.
- Luders, E., Toga, A.W., Thompson, P.M., 2014. Why size matters: differences in brain volume account for apparent sex differences in callosal anatomy: the sexual dimorphism of the corpus callosum. *Neuroimage* 84, 820–824.
- Luo, Z., Hou, C., Wang, L., Hu, D., 2019. Gender identification of human cortical 3-D morphology using hierarchical sparsity. *Front. Hum. Neurosci.* 13, 29.
- Lv, B., Li, J., He, H., Li, M., Zhao, M., Ai, L., Yan, F., Xian, J., Wang, Z., 2010. Gender consistency and difference in healthy adults revealed by cortical thickness. *Neuroimage* 53, 373–382.
- Maeda, Y., Yoon, S., 2013. A meta-analysis on gender differences in mental rotation ability measured by the purdue spatial visualization tests: visualization of rotations (PSVT:R). *Educ. Psychol. Rev.* 25, 69–94.
- Maguire, E.A., Burgess, N., O'Keefe, J., 1999. Human spatial navigation: cognitive maps, sexual dimorphism, and neural substrates. *Curr. Opin. Neurobiol.* 9, 171–177.
- Makowski, C., Béland, S., Kostopoulos, P., Bhagwat, N., Devenyi, G.A., Malla, A.K., Joober, R., Lepage, M., Chakravarty, M.M., 2018. Evaluating accuracy of striatal, pallidal, and thalamic segmentation methods: comparing automated approaches to manual delineation. *Neuroimage* 170, 182–198.
- Maney, D.L., 2015. Just like a circus: the public consumption of sex differences. *Curr. Top. Behav. Neurosci.* 19, 279–296.
- Martinez, K., Janssen, J., Pineda-Pardo, J.A., Carmona, S., Roman, F.J., Aleman-Gomez, Y., García-García, D., Escorial, S., Quiroga, M.A., Santarneccchi, E., Navas-Sánchez, F.J., Descos, M., Arango, C., Colom, R., 2017. Individual differences in the dominance of interhemispheric connections predict cognitive ability beyond sex and brain size. *Neuroimage* 155, 234–244.
- Martínez, K., Madsen, S.K., Joshi, A.A., Joshi, S.H., Roman, F.J., Villalon-Reina, J., Burgaleta, M., Karama, S., Janssen, J., Marinetto, E., Descos, M., Thompson, P.M., Colom, R., 2015. Reproducibility of brain-cognition relationships using three cortical surface-based protocols: an exhaustive analysis based on cortical thickness. *Hum. Brain Map.* 36, 3227–3245.
- Marwha, D., Halari, M., Eliot, L., 2017. Meta-analysis reveals a lack of sexual dimorphism in human amygdala volume. *Neuroimage* 147, 282–294.
- McCarthy, M.M., 2016. Multifaceted origins of sex differences in the brain. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150106.
- McCarthy, M.M., Arnold, A.P., Ball, G.F., Blaustein, J.D., De Vries, G.J., 2012. Sex differences in the brain: the not so inconvenient truth. *J. Neurosci.* 32, 2241–2247.
- McClure, E.B., 2000. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychol. Bull.* 126, 424–453.
- McDowell, M.A., Fryar, C.D., Ogden, C.L., Flegal, K.M., 2008. Anthropometric reference data for children and adults: United States, 2003–2006. *Health Stat. Report.* 1–48.
- McGlove, J., 1980. Sex differences in human brain asymmetry: a critical survey. *Behav. Brain Sci.* 3, 215–263.
- McKay, D.R., Knowles, E.E.M., Winkler, A.A.M., Sprooten, E., Kochunov, P., Olvera, R.L., Curran, J.E., Kent Jr, J.W., Carless, M.A., Göring, H.H.H., Dyer, T.D., Duggirala, R.,

- Almasy, L., Fox, P.T., Blangero, J., Glahn, D.C., Kent Jr, J.W., 2014. Influence of age, sex and genetic factors on the human brain. *Brain Imaging Behav.* 8, 143–152.
- McLean, C.P., Asnaani, A., Litz, B.T., Hofmann, S.G., 2011. Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *J. Psychiatr. Res.* 45, 1027–1035.
- Mechelli, A., Friston, K.J., Frackowiak, R.S., Price, C.J., 2005. Structural covariance in the human cortex. *J. Neurosci.* 25, 8303–8310.
- Mercure, E., Ashwin, E., Dick, F., Halit, H., Auyeung, B., Baron-Cohen, S., Johnson, M.H., 2009. IQ, fetal testosterone and individual variability in children's functional lateralization. *Neuropsychologia* 47, 2537–2543.
- Methil, D., 2019. 6 Ways Men and Women Are Different at the Place of Work, 23-Apr-2019. Science Times. <https://www.sciencetimes.com/articles/20689/20190423/6-ways-men-and-women-are-different-at-the-place-of-work.htm>.
- Mills, K.L., Lalonde, F., Clasen, L.S., Giedd, J.N., Blakemore, S., 2014. Developmental changes in the structure of the social brain in late childhood and adolescence. *Soc. Cogn. Affect. Neurosci.* 9, 123–131.
- Mitracheva, E., Kimura, R., Logothetis, N.K., Noori, H.R., 2019. Neural substrates of sexual arousal are not sex dependent. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15671–15676.
- Morey, R.A., Petty, C.M., Xu, Y., Pannu Hayes, J., Wagner, H.R., Lewis, D.V., LaBar, K.S., Styner, M., McCarthy, G., 2009. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* 45, 855–866.
- Mu, Q., Xie, J., Wen, Z., Weng, Y., Shuyun, Z., 1999. A quantitative MR study of the hippocampal formation, the amygdala, and the temporal horn of the lateral ventricle in healthy subjects 40 to 90 years of age. *AJNR Am. J. Neuroradiol.* 20, 207–211.
- Murray, C.A., 2020. Human Diversity: The Biology of Gender, Race, and Class. Twelve., New York.
- National Institutes of Health, 2015. Consideration of Sex As a Biological Variable in NIH-funded Research. <https://orwh.od.nih.gov/sites/orwh/files/docs/NOT-OD-15-102-Guidance.pdf>.
- Nazareth, A., Huang, X., Voyer, D., Newcombe, N., 2019. A meta-analysis of sex differences in human navigation skills. *Psychon. Bull. Rev.* 26, 1503–1528.
- Neufang, S., Specht, K., Hausmann, M., Gunturkun, O., Herpertz-Dahlmann, B., Fink, G.R., Konrad, K., 2009. Sex differences and the impact of steroid hormones on the developing human brain. *Cereb. Cortex* 19, 464–473.
- Nielsen, J.A., Zielinski, B.A., Ferguson, M.A., Lainhart, J.E., Anderson, J.S., 2013. An evaluation of the left-brain vs. right-brain hypothesis with resting state functional connectivity magnetic resonance imaging. *PLoS One* 8, 1–11.
- Noble, K.G., Grieve, S.M., Korgaonkar, M.S., Engelhardt, L.E., Griffith, E.Y., Williams, L.M., Brickman, A.M., 2012. Hippocampal volume varies with educational attainment across the life-span. *Front. Hum. Neurosci.* 6, 307.
- Nopoulos, P., Flaum, M., O'Leary, D., Andreasen, N.C., 2000. Sexual dimorphism in the human brain: evaluation of tissue volume, tissue composition and surface anatomy using magnetic resonance imaging. *Psychiatry Res. Neuroimaging* 98, 1–13.
- Nopoulos, P.C., Rideout, D., Crespo-Facorro, B., Andreasen, N.C., 2001. Sex differences in the absence of massa intermedia in patients with schizophrenia versus healthy controls. *Schizophr. Res.* 48, 177–185.
- Nota, N.M., Kreukels, B.P.C., den Heijer, M., Veltman, D.J., Cohen-Kettenis, P., Burke, S.M., Bakker, J., 2017. Brain functional connectivity patterns in children and adolescents with gender dysphoria: Sex-atypical or not? *Psychoneuroendocrinology* 86, 187–195.
- Nottebohm, F., Arnold, A.P., 1976. Sexual dimorphism in vocal control areas of the songbird brain. *Science* 194, 211–213.
- O'Brien, L.M., Ziegler, D.A., Deutsch, C.K., Frazier, J.A., Herbert, M.R., Locascio, J.J., 2011. Statistical adjustments for brain size in volumetric neuroimaging studies: some practical implications in methods. *Psychiatry Res. Neuroimaging* 193, 113–122.
- Ohnishi, T., Matsuda, H., Hirakata, M., Ugawa, Y., 2006. Navigation ability dependent neural activation in the human brain: an fMRI study. *Neurosci. Res.* 55, 361–369.
- Pakkenberg, B., Gundersen, H.J., 1997. Neocortical neuron number in humans: effect of sex and age. *J. Comp. Neurol.* 384, 312–320.
- Pallayova, M., Brandeburova, A., Tokarova, D., 2019. Update on sexual dimorphism in brain structure-function interrelationships: a literature review. *Appl. Psychophysiol. Biofeedback* 44, 271–284.
- Patsopoulos, N.A., Tsiatsios, A., Ioannidis, J.P.A., 2007. Claims of sex differences: an empirical assessment in genetic associations. *JAMA* 298, 880–893.
- Paus, T., 2010. Sex differences in the human brain: a developmental perspective. *Prog. Brain Res.* 186, 13–28.
- Paus, T., Wong, A.P., Syme, C., Pausova, Z., 2017. Sex differences in the adolescent brain and body: findings from the Saguenay youth study. *J. Neurosci. Res.* 95, 362–370.
- Pedersen, P.M., Jørgensen, H.S., Nakayama, H., Raaschou, H.O., Olsen, T.S., 1995. Aphasia in acute stroke: incidence, determinants, and recovery. *Ann. Neurol.* 38, 659–666.
- Pedersen, P.M., Vinter, K., Olsen, T.S., 2004. Aphasia after stroke: type, severity and prognosis. The copenhagen aphasia study. *Cerebrovasc. Dis.* 17, 35–43.
- Pell, G.S., Briellmann, R.S., Chan, C.H., Pardoe, H., Abbott, D.F., Jackson, G.D., 2008. Selection of the control group for VBM analysis: influence of covariates, matching and sample size. *Neuroimage* 41, 1324–1335.
- Perlaki, G., Horvath, R., Nagy, S.A., Bogner, P., Doczi, T., Janszky, J., Orsi, G., 2017. Comparison of accuracy between FSL's FIRST and freesurfer for caudate nucleus and putamen segmentation. *Sci. Rep.* 7, 2418.
- Perrin, J.S., Leonard, G., Perron, M., Pike, G.B., Pitiot, A., Richer, L., Veillette, S., Pausova, Z., Paus, T., 2009. Sex differences in the growth of white matter during adolescence. *Neuroimage* 45, 1055–1066.
- Perrone-Bertolotti, M., Pichat, C., Le Bas, J.F., Baciu, A., Baciu, M., 2011. Functional MRI evidence for modulation of cerebral activity by grapheme-to-phoneme conversion in French, and by the variable of gender. *J. Neurolinguistics* 24, 507–520.
- Persson, J., Herlitz, A., Engman, J., Morell, A., Sjölie, D., Wikström, J., Söderlund, H., 2013. Remembering our origin: gender differences in spatial memory are reflected in gender differences in hippocampal lateralization. *Behav. Brain Res.* 256, 219–228.
- Peters, M., Jancke, L., Staiger, J.F., Schlaug, G., Huang, Y., Steinmetz, H., 1998. Unsolved problems in comparing brain sizes in homo sapiens. *Brain Cognit.* 37, 254–285.
- Pfannkuche, K.A., Bouma, A., Groothuis, T.G.G., 2009. Does testosterone affect lateralization of brain and behaviour? A meta-analysis in humans and other animal species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 929–942.
- Pintzka, C.W.S., Hansen, T.I., Evensmoen, H.R., Häger, A.K., 2015. Marked effects of intracranial volume correction methods on sex differences in neuroanatomical structures: a HUNT MRI study. *Front. Neurosci.* 9, 238.
- Plante, E., Schmithorst, V.J., Holland, S.K., Byars, A.W., 2006. Sex differences in the activation of language cortex during childhood. *Neuropsychologia* 44, 1210–1221.
- Plowman, E., Hentz, B., Ellis, C., 2012. Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. *J. Eval. Clin. Pract.* 18, 689–694.
- Polanczyk, G., De Lima, M.S., Horta, B.L., Biederman, J., Rohde, L.A., 2007. The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *Am. J. Psychiatry* 164, 942–948.
- Potvin, O., Mouihha, A., Dieumegarde, L., Duchesne, S., Alzheimer's Disease Neuroimaging Initiative, 2016. Normative data for subcortical regional volumes over the lifetime of the adult human brain. *Neuroimage* 137, 9–20.
- Potvin, O., Dieumegarde, L., Duchesne, S., 2017. Normative morphometric data for cerebral cortical areas over the lifetime of the adult human brain. *Neuroimage* 156, 315–339.
- Potvin, O., Dieumegarde, L., Duchesne, S., 2018a. Corrigendum to 'normative morphometric data for cerebral cortical areas over the lifetime of the adult human brain' [NeuroImage 156 (2017) 315–339]. *Neuroimage* 183, 996–998.
- Potvin, O., Mouihha, A., Dieumegarde, L., Duchesne, S., 2018b. Corrigendum to 'normative data for subcortical regional volumes over the lifetime of the adult human brain' [NeuroImage 137 (2016) 9–20]. *Neuroimage* 183, 994–995.
- Purves, D., Augustine, G., Fitzpatrick, D., Hall, W.C., LaMantia, A., Mooney, R., White, L.E., 2018. Neuroscience, 6th ed. Sinauer, New York.
- Rabinowicz, T., Petetot, J.M., Gartside, P.S., Sheyn, D., Sheyn, T., de, C., 2002. Structure of the cerebral cortex in men and women. *J. Neuropathol. Exp. Neurol.* 61, 46–57.
- Rajamanickam, K., 2020. A mini review on different methods of functional-MRI data analysis. *Arch Intern Med Res* 3, 44–60.
- Rawlik, K., Canela-Xandri, O., Tenesa, A., 2016. Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome Biol.* 17 (166-x).
- Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K.M., Williamson, A., Acker, J.D., 2004. Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. *Neurobiol. Aging* 25, 377.
- Raznahan, A., Lee, Y., Stidd, R., Long, R., Greenstein, D., Clasen, L., Addington, A., Goftay, N., Rapoport, J.L., Giedd, J.N., 2010. Longitudinally mapping the influence of sex and androgen signaling on the dynamics of human cortical maturation in adolescence. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16988–16993.
- Reardon, P.K., Clasen, L., Giedd, J.N., Blumenthal, J., Lerch, J.P., Chakravarty, M.M., Raznahan, A., 2016. An allometric analysis of sex and sex chromosome dosage effects on subcortical anatomy in humans. *J. Neurosci.* 36, 2438–2448.
- Redolfi, A., Manset, D., Barkhof, F., Wahlund, L., Glatard, T., Mangin, J., Frisoni, G.B., neuGRID Consortium, Alzheimer's Disease Neuroimaging Initiative, 2015. Head-to-head comparison of two popular cortical thickness extraction algorithms: a cross-sectional and longitudinal study. *PLoS One* 10, 3.
- Reid, A.T., Hoffstaedter, F., Gong, G., Laird, A.R., Fox, P., Evans, A.C., Amunts, K., Eickhoff, S.B., 2017. A seed-based cross-modal comparison of brain connectivity measures. *Brain Struct. Funct.* 222, 1131–1151.
- Rijpkema, M., Everaerd, D., van, d P., Franke, B., Tendolkar, I., Fernández, G., 2012. Normal sexual dimorphism in the human basal ganglia. *Hum. Brain Mapp.* 33, 1246–1252.
- Ringo, J.L., Doty, R.W., Demeter, S., Simard, P.Y., 1994. Time is of the essence: a conjecture that hemispheric specialization arises from interhemispheric conduction delay. *Cereb. Cortex* 4, 331–343.
- Rippon, G., 2016. The trouble with girls? *Psychologist (United Kingdom)* 29, 918–922.
- Ritchie, S.J., Cox, S.R., Shen, X., Lombardo, M.V., Reus, L.M., Alloza, C., Harris, M.A., Alderson, H.L., Hunter, S., Neilson, E., 2018. Sex differences in the adult human brain: evidence from 5216 UK Biobank participants. *Cereb. Cortex* 28, 2959–2975.
- Roberts, J.E., Bell, M.A., 2000. Sex differences on a mental rotation task: variations in electroencephalogram hemispheric activation between children and college students. *Dev. Neuropsychol.* 17, 199–223.
- Roberts, J.E., Bell, M.A., 2003. Two- and three-dimensional mental rotation tasks lead to different parietal laterality for men and women. *Int. J. Psychophysiol.* 50, 235–246.
- Rosenthal, R., 1979. The file drawer problem and tolerance for null results. *Psychol. Bull.* 86, 638–641.
- Rubia, K., Hyde, Z., Halarie, R., Giampietro, V., Smith, A., 2010. Effects of age and sex on developmental neural networks of visual-spatial attention allocation. *Neuroimage* 51, 817–827.
- Ruigrok, A.N.V., Salimi-Khorshidi, G., Lai, M., Baron-Cohen, S., Lombardo, M.V., Tait, R.J., Suckling, J., 2014. A meta-analysis of sex differences in human brain structure. *Neurosci. Biobehav. Rev.* 39, 34–50.
- Rumberg, B., Baars, A., Fiebach, J., Ladd, M.E., Forsting, M., Senf, W., Gizewski, E.R., 2010. Cycle and gender-specific cerebral activation during a verb generation task using fMRI: comparison of women in different cycle phases, under oral contraception, and men. *Neurosci. Res.* 66, 366–371.

- Rutter, M., Caspi, A., Fergusson, D., Horwood, L.J., Goodman, R., Maughan, B., Moffitt, T.E., Meltzer, H., Carroll, J., 2004. Sex differences in developmental reading disability: new findings from 4 epidemiological studies. *JAMA* 291, 2007–2012.
- Sacher, J., Neumann, J., Okon-Singer, H., Gotowicz, S., Villringer, A., 2013. Sexual dimorphism in the human brain: evidence from neuroimaging. *Magn. Reson. Imaging* 31, 366–375.
- Salat, D.H., Buckner, R.L., Snyder, A.Z., Greve, D.N., Desikan, R.S.R., Busa, E., Morris, J.C., Dale, A.M., Fischl, B., 2004. Thinning of the cerebral cortex in aging. *Cereb. Cortex* 14, 721–730.
- Sale, M.V., Reid, L.B., Cocchi, L., Pagnozzi, A.M., Rose, S.E., Mattingley, J.B., 2017. Brain changes following four weeks of unimanual motor training: evidence from behavior, neural stimulation, cortical thickness, and functional MRI. *Hum. Brain Mapp.* 38, 4773–4787.
- Salk, R.H., Hyde, J.S., Abramson, L.Y., 2017. Gender differences in depression in representative national samples: meta-analyses of diagnoses and symptoms. *Psychol. Bull.* 143, 783–822.
- Sanchis-Segura, C., Ibañez-Gual, M.V., Adrian-Ventura, J., Aguirre, N., Gomez-Cruz, A.J., Avila, C., Forn, C., 2019. Sex differences in gray matter volume: How many and how large are they really? *Biol. Sex Diff.* 10, 32.
- Sanchis-Segura, C., Ibañez-Gual, M.V., Aguirre, N., Cruz-Gómez, Á.J., Forn, C., 2020. Effects of different intracranial volume correction methods on univariate sex differences in grey matter volume and multivariate sex prediction. *Sci. Rep.* 10, 1–15.
- Sanfilipo, M.P., Benedict, R.H.B., Zivadinov, R., Bakshi, R., 2004. Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. *Neuroimage* 22, 1732–1743.
- Satterthwaite, T.D., Wolf, D.H., Roalf, D.R., Ruparel, K., Erus, G., Vandekar, S., Gennatas, E.D., Elliott, M.A., Smith, A., Hakanson, H., Verma, R., Davatzikos, C., Gur, R.E., Gur, R.C., 2015. Linked sex differences in cognition and functional connectivity in youth. *Cereb. Cortex* 25, 2383–2394.
- Scheinost, D., Finn, E.S., Tokoglu, F., Shen, X., Papademetris, X., Hampson, M., Constable, R.T., 2015. Sex differences in normal age trajectories of functional brain networks. *Hum. Brain Mapp.* 36, 1524–1535.
- Schlosser, R., Hutchinson, M., Joseffer, S., Rusinek, H., Saarimaki, A., Stevenson, J., Dewey, S.L., Brodie, J.D., 1998. Functional magnetic resonance imaging of human brain activity in a verbal fluency task. *J. Neurol. Neurosurg. Psychiatry* 64, 492–498.
- Schmidt, I.M., Mølgaard, C., Main, K.M., Michaelsen, K.F., 2001. Effect of gender and lean body mass on kidney size in healthy 10-year-old children. *Pediatr. Nephrol.* 16, 366–370.
- Schmidt, H., Jogia, J., Fast, K., Christodoulou, T., Haldane, M., Kumari, V., Frangou, S., 2009. No gender differences in brain activation during the N-back task: an fMRI study in healthy individuals. *Hum. Brain Mapp.* 30, 3609–3615.
- Schmithorst, V.J., Holland, S.K., 2006. Functional MRI evidence for disparate developmental processes underlying intelligence in boys and girls. *Neuroimage* 31, 1366–1379.
- Schoemaker, D., Buss, C., Head, K., Sandman, C.A., Davis, E.P., Chakravarty, M.M., Gauthier, S., Pruessner, J.C., 2016. Hippocampus and amygdala volumes from magnetic resonance images in children: assessing accuracy of FreeSurfer and FSL against manual segmentation. *Neuroimage* 129, 1–14.
- Schöning, S., Engelien, A., Kugel, H., Schafer, S., Schiffbauer, H., Zwitserlood, P., Pletziger, E., Beizai, P., Kersting, A., Ohrmann, P., Greb, R.R., Lehmann, W., Heindel, W., Arolt, V., Konrad, C., 2007. Functional anatomy of visuo-spatial working memory during mental rotation is influenced by sex, menstrual cycle, and sex steroid hormones. *Neuropsychologia* 45, 3203–3214.
- Schweinsburg, A.D., Nagel, B.J., Tapert, S.F., 2005. fMRI reveals alteration of spatial working memory networks across adolescence. *J. Int. Neuropsychol. Soc.* 11, 631–644.
- Semrud-Clikeman, M., Fine, J.G., Bledsoe, J., Zhu, D.C., 2012. Gender differences in brain activation on a mental rotation task. *Int. J. Neurosci.* 122, 590–597.
- Sen, F., Ulubay, H., Ozeksi, P., Sargon, M.F., Tascioglu, A.B., 2005. Morphometric measurements of the thalamus and interthalamic adhesion by MR imaging. *Neuroanatomy* 4, 10–12.
- Sepehrband, F., Lynch, K.M., Cabeen, R.P., Gonzalez-Zacarias, C., Zhao, L., D'arcy, M., Kesselman, C., Herting, M.M., Dinov, I.D., Toga, A.W., 2018. Neuroanatomical morphometric characterization of sex differences in youth using statistical learning. *Neuroimage* 172, 217–227.
- Sergerie, K., Chochol, C., Armony, J.L., 2008. The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 32, 811–830.
- Seurinck, R., Vingerhoets, G., De Lange, F.P., Achten, E., 2004. Does egocentric mental rotation elicit sex differences? *Neuroimage* 23, 1440–1449.
- Shahab, S., Stefanik, L., Foussias, G., Lai, M., Anderson, K.K., Voineskos, A.N., 2018. Sex and diffusion tensor imaging of white matter in schizophrenia: a systematic review plus meta-analysis of the corpus callosum. *Schizophr. Bull.* 44, 203–221.
- Shaywitz, B.A., Shaywitz, S.E., Pugh, K.R., Constable, R.T., Skudlarski, P., Fulbright, R.K., Bronen, R.A., Fletcher, J.M., Shankweiler, D.P., Katz, L., 1995. Sex differences in the functional organization of the brain for language. *Nature* 373, 607–609.
- Shiino, A., Chen, Y., Tanigaki, K., Yamada, A., Vigers, P., Watanabe, T., Tooyama, I., Akiguchi, I., 2017. Sex-related difference in human white matter volumes studied: inspection of the corpus callosum and other white matter by VBM. *Sci. Rep.* 7, 1–7.
- Siddaway, A.P., Wood, A.M., Hedges, L.V., 2019. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Ann. Rev. Psychol.* 70, 747–770.
- Smith, R.J., 2005. Relative size versus controlling for size: interpretation of ratios in research on sexual dimorphism in the human corpus callosum. *Curr. Anthropol.* 46, 249–273.
- Smith, D.V., Utevsky, A.V., Bland, A.R., Clement, N., Clithero, J.A., Harsch, A.E., Carter, R.M., Huettel, S.A., 2014. Characterizing individual differences in functional connectivity using dual-regression and seed-based approaches. *Neuroimage* 95, 1–12.
- Smith, E.S., Junger, J., Derntl, B., Habel, U., 2015. The transsexual brain—A review of findings on the neural basis of transsexualism. *Neurosci. Biobehav. Rev.* 59, 251–266.
- Sneider, J.T., Rogowska, J., Sava, S., Yurgelun-Todd, D.A., 2011. A preliminary study of sex differences in brain activation during a spatial navigation task in healthy adults. *Percept. Mot. Skills* 113, 461–480.
- Snyder, P.J., Bogerts, B., Wu, H., Bilder, R.M., Deoras, K.S., Lieberman, J.A., 1998. Absence of the adhesio interthalamic as a marker of early developmental neuropathology in schizophrenia: an MRI and postmortem histologic study. *J. Neuroimaging* 8, 159–163.
- Solé-Padellés, C., Castro-Fornieles, J., de la Serna, E., Calvo, R., Baeza, I., Moya, J., Lázaro, Rosa M., Bargalló, N., Sugranyes, G., 2016. Intrinsic connectivity networks from childhood to late adolescence: effects of age and sex. *Dev Cog Neurosci* 17, 35–44.
- Sommer, I.E., Aleman, A., Bouma, A., Kahn, R.S., 2004. Do women really have more bilateral language representation than men? A meta-analysis of functional imaging studies. *Brain* 127, 1845–1852.
- Sommer, I.E., Aleman, A., Somers, M., Boks, M.P., Kahn, R.S., 2008. Sex differences in handedness, asymmetry of the planum temporale and functional language lateralization. *Brain Res.* 1206, 76–88.
- Song, F., Parekh, S., Hooper, L., Luke, Y., Ryder, J., Sutton, A., Hing, C., Kwok, C., Pang, C., Harvey, I., 2010. Dissemination and publication of research findings: an updated review of related biases. *Health Technol. Assess.* 14, 1–220.
- Sowell, E.R., Trauner, D.A., Gamst, A., Jernigan, T.L., 2002. Development of cortical and subcortical brain structures in childhood and adolescence: a structural MRI study. *Dev. Med. Child Neurol.* 44, 4–16.
- Sowell, E.R., Peterson, B.S., Kan, E., Woods, R.P., Yoshii, J., Bansal, R., Xu, D., Zhu, H., Thompson, P.M., Toga, A.W., 2007. Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. *Cereb. Cortex* 17, 1550–1560.
- Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., Chang, L., 2000. Gender differences in the functional organization of the brain for working memory. *Neuroreport* 11, 2581–2585.
- Springer, K.W., Mager Stellman, J., Jordan-Young, R., 2012. Beyond a catalogue of differences: a theoretical frame and good practice guidelines for researching s/g in human health. *Soc. Sci. Med.* 74, 1817–1824.
- Squeglia, L.M., Schweinsburg, A.D., Pulido, C., Tapert, S.F., 2011. Adolescent binge drinking linked to abnormal spatial working memory brain activation: differential gender effects. *Alcohol. Clin. Exp. Res.* 35, 1831–1841.
- Staphorius, A.S., Kreukels, B.P., Cohen-Kettenis, P.T., Veltman, D.J., Burke, S.M., Schagen, S.E., Wouters, F.M., Delemarre-van de Waal, H.A., Bakker, J., 2015. Puberty suppression and executive functioning: an fMRI-study in adolescents with gender dysphoria. *Psychoneuroendocrinology* 56, 190–199.
- Stevens, J.S., Hamann, S., 2012. Sex differences in brain activation to emotional stimuli: a meta-analysis of neuroimaging studies. *Neuropsychologia* 50, 1578–1593.
- Sullivan, E.V., Pfefferbaum, A., Swan, G.E., Carmelli, D., 2001. Heritability of hippocampal size in elderly twin men: equivalent influence from genes and environment. *Hippocampus* 11, 754–762.
- Sun, Y., Lee, R., Chen, Y., Collinson, S., Thakor, N., Bezerianos, A., Sim, K., 2015. Progressive gender differences of structural brain networks in healthy adults: a longitudinal, diffusion tensor imaging study. *PLoS One* 10, 1–18.
- Suo, C., Leon, I., Brodaty, H., Trollor, J., Wen, W., Sachdev, P., Valenzuela, M.J., 2012. Supervisory experience at work is linked to low rate of hippocampal atrophy in late life. *Neuroimage* 63, 1542–1551.
- Szucs, D., Ioannidis, J.P., 2017. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biol.* 15, e2000797.
- Takahashi, T., Suzuki, M., Zhou, S., Nakamura, K., Tanino, R., Kawasaki, Y., Seal, M.L., Seto, H., Kurachi, M., 2008a. Prevalence and length of the adhesio interthalamic in schizophrenia spectrum disorders. *Psychiatry Res. Neuroimaging* 164, 90–94.
- Takahashi, T., Yücel, M., Yung, A.R., Wood, S.J., Phillips, L.J., Berger, G.E., Ang, A., Soulsby, B., McGorry, P.D., Suzuki, M., Velakoulis, D., Pantelis, C., 2008b. Adhesio interthalamic in individuals at high-risk for developing psychosis and patients with psychotic disorders. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 32, 1708–1714.
- Takao, H., Hayashi, N., Ohtomo, K., 2014. Sex dimorphism in the white matter: fractional anisotropy and brain size. *J. Magn. Reson. Imaging* 39, 917–923.
- Talge, N.M., Mudd, L.M., Sikorskii, A., Basso, O., 2014. United states birth weight reference corrected for implausible gestational age estimates. *Pediatrics* 133, 844–853.
- Tan, A., Ma, W., Vira, A., Marwha, D., Eliot, L., 2016. The human hippocampus is not sexually-dimorphic: meta-analysis of structural MRI volumes. *Neuroimage* 124, 350–366.
- Tang, T., Jiao, Y., Wang, X., Lu, Z., 2013. Gender versus brain size effects on subcortical gray matter volumes in the human brain. *Neurosci. Lett.* 556, 79–83.
- Tanner, J.M., 1990. *Foetus into Man: Physical Growth From Conception to Maturity*. Harvard University Press, Cambridge, Mass.
- Thompson, A.E., Voyer, D., 2014. Sex differences in the ability to recognise non-verbal displays of emotion: a meta-analysis. *Cogn. Emot.* 28, 1164–1195.
- Tian, L., Wang, J., Yan, C., He, Y., 2011. Hemisphere- and gender-related differences in small-world brain networks: a resting-state functional MRI study. *Neuroimage* 54, 191–202.

- Tomasi, D., Volkow, N.D., 2012a. Gender differences in brain functional connectivity density. *Hum. Brain Mapp.* 33, 849.
- Tomasi, D., Volkow, N.D., 2012b. Laterality patterns of brain functional connectivity: gender effects. *Cereb. Cortex* 22, 1455–1462.
- Trzesniak, C., Kempton, M.J., Busatto, G.F., Oliveira, I.Rd, Galvão-de Almeida, A., Kambeitz, J., Ferrari, M.C.F., Filho, A.S., Chagas, M.H.N., Zuardi, A.W., Hallak, J.E.C., McGuire, P.K., Crippa, J.A.S., 2011. Adhesio interthalamic alterations in schizophrenia spectrum disorders: a systematic review and meta-analysis. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 877–886.
- Tschernegg, M., Neuper, C., Schmidt, R., Wood, G., Kronbichler, M., Fazekas, F., Enzinger, C., Koenig, M., 2017. fMRI to probe sex-related differences in brain function with multitasking. *PLoS One* 12, e0181554.
- Tunç, B., Solmaz, B., Parker, D., Satterthwaite, T.D., Elliott, M.A., Calkins, M.E., Ruparel, K., Gur, R.E., Gur, R.C., Verma, R., 2016. Establishing a link between sex-related differences in the structural connectome and behaviour. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150111.
- Unterrainer, J.M., Ruff, C.C., Rahm, B., Kaller, C.P., Spreer, J., Schwarzwald, R., Halsband, U., 2005. The influence of sex differences and individual task performance on brain activation during planning. *Neuroimage* 24, 586–590.
- Valera, E.M., Brown, A., Biederman, J., Faraone, S.V., Makris, N., Monuteaux, M.C., Whitfield-Gabrieli, S., Vitaliano, M., Schiller, M., Seidman, L.J., 2010. Sex differences in the functional neuroanatomy of working memory in adults with ADHD. *Am. J. Psychiatry* 167, 86–94.
- van Hemmen, J., Veltman, D.J., Hoekzema, E., Cohen-Kettenis, P., Dessens, A.B., Bakker, J., 2016. Neural activation during mental rotation in complete androgen insensitivity syndrome: the influence of sex hormones and sex chromosomes. *Cereb. Cortex* 26, 1036–1045.
- van Putten, M.J., Olbrich, S., Arns, M., 2018. Predicting sex from brain rhythms with deep learning. *Sci. Rep.* 8, 1–7.
- Verde, P., Piccardi, L., Bianchini, F., Trivelloni, P., Guariglia, C., Tomao, E., 2013. Gender effects on mental rotation in pilots vs. nonpilots. *Aviat. Space Environ. Med.* 84, 726.
- Verde, P., Piccardi, L., Bianchini, F., Guariglia, C., Carrozzo, P., Morgagni, F., Boccia, M., 2015. Gender differences in navigational memory: pilots vs. nonpilots. *Aerospace Med. Hum. Perform.* 86, 103.
- Vijayakumar, N., Allen, N.B., Youssef, G., Dennison, M., Yücel, M., Simmons, J.G., Whittle, S., 2016. Brain development during adolescence: a mixed-longitudinal investigation of cortical thickness, surface area, and volume. *Hum. Brain Mapp.* 37, 2027–2038.
- Voevodskaya, O., Simmons, A., Nordenskjöld, R., Kullberg, J., Ahlström, H., Lind, L., Wahlund, L., Larsson, E., Westman, E., 2014. The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Front. Aging Neurosci.* 6, 264.
- Voyer, D., 1996. On the magnitude of laterality effects and sex differences in functional lateralities. *Laterality Asymmetries Body Brain Cogn.* 1, 51–84.
- Voyer, D., Voyer, S., Bryden, M.P., 1995. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol. Bull.* 117, 250–270.
- Wager, T.D., Phan, K.L., Liberzon, I., Taylor, S.F., 2003. Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage* 19, 513–531.
- Wallach, J.D., Sullivan, P.G., Trepanowski, J.F., Steyerberg, E.W., Ioannidis, J.P.A., 2016. Sex based subgroup differences in randomized controlled trials: empirical evidence from cochrane meta-analyses. *BMJ* 355, 15826.
- Wallentin, M., 2009. Putative sex differences in verbal abilities and language cortex: a critical review. *Brain Lang.* 108, 175–183.
- Wang, L., Shen, H., Tang, F., Zang, Y., Hu, D., 2012. Combined structural and resting-state functional MRI analysis of sexual dimorphism in the young adult human brain: an MVPA approach. *Neuroimage* 61, 931–940.
- Watila, M.M., Balarabe, S.A., 2015. Factors predicting post-stroke aphasia recovery. *J. Neurol. Sci.* 352, 12–18.
- Wei, W., Chen, C., Dong, Q., Zhou, X., 2016. Sex differences in gray matter volume of the right anterior hippocampus explain sex differences in three-dimensional mental rotation. *Front. Hum. Neurosci.* 10, 580.
- Weis, S., Hausmann, M., Stoffers, B., Vohn, R., Kellermann, T., Sturm, W., 2008. Estradiol modulates functional brain organization during the menstrual cycle: an analysis of interhemispheric inhibition. *J. Neurosci.* 28, 13401–13410.
- Weiss, E.M., Siedentopf, C., Hofer, A., Deisenhammer, E.A., Hoptman, M.J., Kremser, C., Golaszewski, S., Felber, S., Fleischhacker, W.W., Delazer, M., 2003a. Brain activation pattern during a verbal fluency test in healthy male and female volunteers: a functional magnetic resonance imaging study. *Neurosci. Lett.* 352, 191–194.
- Weiss, E., Siedentopf, C.M., Hofer, A., Deisenhammer, E.A., Hoptman, M.J., Kremser, C., Golaszewski, S., Felber, S., Fleischhacker, W.W., Delazer, M., 2003b. Sex differences in brain activation pattern during a visuospatial cognitive task: a functional magnetic resonance imaging study in healthy volunteers. *Neurosci. Lett.* 344, 169–172.
- Weissman-Fogel, I., Moayed, M., Taylor, K.S., Pope, G., Davis, K.D., 2010. Cognitive and default-mode resting state networks: Do male and female brains' rest differently? *Hum. Brain Mapp.* 31, 1713–1726.
- Westerhausen, R., Fjell, A.M., Krogsgaard, S.K., Rohani, D.A., Skrane, J.S., Häberg, A.K., Walhovd, K.B., 2016. Selective increase in posterior corpus callosum thickness between the age of 4 and 11 years. *Neuroimage* 139, 17–25.
- Whittle, S., Lichten, R., Dennison, M., Vijayakumar, N., Schwartz, O., Byrne, M.L., Simmons, J.G., Yucel, M., Pantelis, C., McGorry, P., Allen, N.B., 2014. Structural brain development and depression onset during adolescence: a prospective longitudinal study. *Am. J. Psychiatry* 171, 564–571.
- Whitwell, J.L., Crum, W.R., Watt, H.C., Fox, N.C., 2001. Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *AJR Am. J. Neuroradiol.* 22, 1483–1489.
- Wierenga, L.M., Langen, M., Oranje, B., Durston, S., 2014. Unique developmental trajectories of cortical thickness and surface area. *Neuroimage* 87, 120–126.
- Witelson, S.F., Nowakowski, R.S., 1991. Left out axons make men right: a hypothesis for the origin of handedness and functional asymmetry. *Neuropsychologia* 29, 327–333.
- Witelson, S.F., Glezer, I.I., Kigar, D.L., 1995. Women have greater density of neurons in posterior temporal cortex. *J. Neurosci.* 15, 3418–3428.
- Witelson, S.F., Beresh, H., Kigar, D.L., 2006. Intelligence and brain size in 100 postmortem brains: sex, lateralization and age factors. *Brain* 129, 386–398.
- Woodruff, T.K., Kibbe, M.R., Paller, A.S., Turek, F.W., Woolley, C.S., 2014. Commentary: "Leaning in" to support sex differences in basic science and clinical research. *Endocrinology* 155, 1181–1183.
- Woollett, K., Maguire, E.A., 2011. Acquiring "the knowledge" of London's layout drives structural brain changes. *Curr. Biol.* 21, 2109–2114.
- Wu, K., Taki, Y., Sato, K., Hashizume, H., Sassa, Y., Takeuchi, H., Thyreau, B., He, Y., Alan, C.E., Li, X., Kawashima, R., Fukuda, H., 2013. Topological organization of functional brain networks in healthy children: differences in relation to age, sex, and intelligence. *PLoS One* 8, e55347.
- Xin, J., Zhang, Y., Tang, Y., Yang, Y., 2019. Brain differences between men and women: evidence from deep learning. *Front. Neurosci.* 13, 185.
- Xu, J., Kobayashi, S., Yamaguchi, S., Iijima, K., Okada, K., Yamashita, K., 2000. Gender effects on age-related changes in brain structure. *AJR Am. J. Neuroradiol.* 21, 112–118.
- Yan, C., Gong, G., Wang, J., Wang, D., Liu, D., Zhu, C., Chen, Z.J., Evans, A., Zang, Y., He, Y., 2011. Sex- and brain size-related small-world structural cortical networks in young adults: a DTI tractography study. *Cereb. Cortex* 21, 449–458.
- Yan, C., Craddock, R.C., Zuo, X., Zang, Y., Milham, M.P., 2013. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. *Neuroimage* 80, 246–262.
- Yang, G., Bozek, J., Han, M., Gao, J.H., 2020a. Constructing and evaluating a cortical surface atlas and analyzing cortical sex differences in young Chinese adults. *Hum. Brain Mapp.* 41, 2495–2513.
- Yang, G., Zhou, S., Bozek, J., Dong, H.M., Han, M., Zuo, X.N., Liu, H., Gao, J.H., 2020b. Sample sizes and population differences in brain template construction. *NeuroImage* 206, 116318.
- Yuan, L., Kong, F., Luo, Y., Zeng, S., Lan, J., You, X., 2019. Gender differences in large-scale and small-scale spatial ability: a systematic review based on behavioral and neuroimaging research. *Front. Behav. Neurosci.* 13, 128.
- Zacks, J.M., 2008. Neuroimaging studies of mental rotation: a meta-analysis and review. *J. Cogn. Neurosci.* 20, 1–19.
- Zell, E., Krizan, Z., Teeter, S.R., 2015. Evaluating gender similarities and differences using metasynthesis. *Am. Psychol.* 70, 10–20.
- Zhang, K., Sejnowski, T.J., 2000. A universal scaling law between gray matter and white matter of cerebral cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5621.
- Zhang, C., Cahill, N.D., Arbabshirani, M.R., White, T., Baum, S.A., Michael, A.M., 2016. Sex and age effects of functional connectivity in early adulthood. *Brain Connect.* 6, 700–713.
- Zhang, C., Dougherty, C.C., Baum, S.A., White, T., Michael, A.M., 2018. Functional connectivity predicts gender: evidence for gender differences in resting brain connectivity. *Hum. Brain Mapp.* 39, 1765–1776.
- Zhao, T., Cao, M., Niu, H., Zuo, X., Evans, A., He, Y., Dong, Q., Shu, N., 2015. Age-related changes in the topological organization of the white matter structural connectome across the human lifespan. *Hum. Brain Mapp.* 36, 3777–3792.
- Zielinski, B.A., Gennatas, E.D., Zhou, J., Seeley, W.W., 2010. Network-level structural covariance in the developing brain. *Proc. Natl. Acad. Sci. U. S. A.* 107, 18191–18196.
- Zilles, K., Kawashima, R., Dabringhaus, A., Fukuda, H., Schormann, T., 2001. Hemispheric shape of European and Japanese brains: 3-D MRI analysis of intersubject variability, ethnical, and gender differences. *NeuroImage* 13, 262–271.
- Zubiaurre-Elorza, L., Junque, C., Gómez-Gil, E., Segovia, S., Carrillo, B., Rametti, G., Guillamon, A., 2013. Cortical thickness in untreated transsexuals. *Cereb. Cortex* 23, 2855–2862.
- Zuo, X., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F.X., Sporns, O., Milham, M.P., 2012. Network centrality in the human functional connectome. *Cereb. Cortex* 22, 1862–1875.