Sex beyond the genitalia: The human brain mosaic

Daphna Joel[a,b,1], Zohar Berman[b], Ido Tavor[c], Nadav Wexler[d], Olga Gaber[a], Yaniv Stein[d], Nisan Shefi[a,b], Jared Pool[e], Sebastian Urchs[e], Daniel S. Margulies[e], Franziskus Liem[e,f], Jürgen Hänggi[f], Lutz Jäncke[f], and Yaniv Assaf[b,c]

[a]School of Psychological Sciences, Tel-Aviv University, Ramat Aviv, Tel-Aviv 6997801, Israel; [b]Sagol School of Neuroscience, Tel-Aviv University, Ramat Aviv, Tel-Aviv 6997801, Israel; [c]Department of Neurobiology, Faculty of Life Sciences, Tel-Aviv University, Ramat Aviv, Tel-Aviv 6997801, Israel; [d]School of Mathematical Sciences, Tel-Aviv University, Ramat Aviv, Tel-Aviv 6997801, Israel; [e]Max Planck Research Group for Neuroanatomy & Connectivity, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany; and [f]Division Neuropsychology, Department of Psychology, University of Zurich, 8050 Zurich, Switzerland

Edited by Bruce S. McEwen, The Rockefeller University, New York, NY, and approved October 23, 2015 (received for review June 4, 2015)

Whereas a categorical difference in the genitals has always been acknowledged, the question of how far these categories extend into human biology is still not resolved. Documented sex/gender differences in the brain are often taken as support of a sexually dimorphic view of human brains ("female brain" or "male brain"). However, such a distinction would be possible only if sex/gender differences in brain features were highly dimorphic (i.e., little overlap between the forms of these features in males and females) and internally consistent (i.e., a brain has only "male" or only "female" features). Here, analysis of MRIs of more than 1,400 human brains from four datasets reveals extensive overlap between the distributions of females and males for all gray matter, white matter, and connections assessed. Moreover, analyses of internal consistency reveal that brains with features that are consistently at one end of the "maleness-femaleness" continuum are rare. Rather, most brains are comprised of unique "mosaics" of features, some more common in females compared with males, some more common in males compared with females, and some common in both females and males. Our findings are robust across sample, age, type of MRI, and method of analysis. These findings are corroborated by a similar analysis of personality traits, attitudes, interests, and behaviors of more than 5,500 individuals, which reveals that internal consistency is extremely rare. Our study demonstrates that, although there are sex/gender differences in the brain, human brains do not belong to one of two distinct categories: male brain/female brain.

gender differences | sex differences | brain structure | brain connectivity | behavior

The question of whether males and females form two distinct categories has attracted thinkers from ancient times to this day. Whereas a categorical difference in the genitals has always been acknowledged, the question of how far these categories extend into human biology is still not resolved (for a historical overview, see refs. 1 and 2). Documented sex/gender* differences in the brain are often taken as support of a sexually dimorphic view of human brains ("female brain" vs. "male brain"), and consequently, of a sexually dimorphic view of human behavior, cognition, personality, attitudes, and other gender characteristics (3). Joel (4, 5) has recently argued that the existence of sex/gender differences in the brain is not sufficient to conclude that human brains belong to two distinct categories. Rather, such a distinction requires the fulfillment of two conditions: one, the form of the elements that show sex/gender differences should be dimorphic, that is, with little overlap between the forms of the elements in males and females. Two, there should be a high degree of internal consistency in the form of the different elements of a single brain (e.g., all elements have the "male" form).

Previous criticisms of the dichotomous view of human brain have focused on the fact that most sex/gender differences are nondimorphic population-level differences with extensive overlap of the distributions of females and males and have therefore claimed that human brains cannot be sorted into two distinct classes: "male brains" and "female brains" (6–8). However, if brains are internally consistent in the degree of "maleness-femaleness" of each of their elements, it will still be possible to align brains on a "male-brain–female-brain" continuum (4, 5). Such an alignment may be predicted by the classic view of sexual differentiation of the brain, according to which masculinization and defeminization of the brain are under the sole influence of testosterone (9). In contrast, more recent evidence that masculinization and feminization are independent processes and that sexual differentiation progresses independently in different brain tissues (10), predicts poor internal consistency (4, 5). Poor internal consistency is further predicted by evidence that the effects of sex may be different and even opposite under different environmental conditions and that these sex-by-environment interactions may be different for different brain features (4, 5). There are indeed examples of lack of internal consistency within a single brain in the animal literature (4, 5), yet it is not clear whether this is a common phenomenon that involves most features that show sex differences and is seen in most individuals. Here we assess the degree of internal consistency in the human brain using data obtained from MRI, a method that allows the simultaneous assessment of multiple brain features in many individuals.

We used datasets obtained from several different imaging modalities and analyzed with different methods to ensure that our conclusion is not measure, analysis, or sample dependent.

Significance

Sex/gender differences in the brain are of high social interest because their presence is typically assumed to prove that humans belong to two distinct categories not only in terms of their genitalia, and thus justify differential treatment of males and females. Here we show that, although there are sex/gender differences in brain and behavior, humans and human brains are comprised of unique "mosaics" of features, some more common in females compared with males, some more common in males compared with females, and some common in both females and males. Our results demonstrate that regardless of the cause of observed sex/gender differences in brain and behavior (nature or nurture), human brains cannot be categorized into two distinct classes: male brain/female brain.

Author contributions: D.J. designed research; D.J., Z.B., S.U., F.L., J.H., and L.J. performed research; D.J. and Y.A. contributed new analytic tools; D.J., Z.B., I.T., N.W., O.G., Y.S., N.S., J.P., S.U., F.L., and J.H. analyzed data; and D.J. and D.S.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Our anonymized raw neuroimaging data and accompanying metadata have been deposited at psy-neuro-nassy.uzh.ch and are accessible with a username and password that can be obtained from the authors by email (djoel@post.tau.ac.il or j.haenggi@psychologie.uzh.ch).

[1]To whom correspondence should be addressed. Email: djoel@post.tau.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1509654112/-/DCSupplemental.

*We use the term sex/gender to indicate that studies typically assess subjects' sex (i.e., whether one is male or female) but observed differences may reflect the effects of both sex and gender (that is, the social construction of sex). We ignore here the important issue of the probable effects of gender on observed differences between females and males in brain and behavior, because we want to emphasize that regardless of the cause of these differences (sex, gender, or their interactions), they do not add up to create two distinct categories, one typical of males and the other typical of females.

The number of subjects in these datasets ranged from 138 to 855. In each dataset, following an assessment of sex/gender differences in all regions, we focused on the regions showing the largest sex/gender differences (i.e., least overlap between females and males). Because also in these regions there was a considerable overlap between the distributions of females and males, which made a division into two distinct forms impossible, we tested whether individuals would be consistently at one end of the "femaleness-maleness" continuum across brain regions or show "substantial variability", being at the one end of the "femaleness-maleness" continuum on some regions and at the other end on other regions. We found that regardless of sample, type of MRI, and method of analysis, substantial variability is much more prevalent than internal consistency.

Table 1. Internal consistency and substantial variability in human brain and behavior

Dataset	Age: range, mean (SD)	Number of characteristics in analysis of internal consistency (number of characteristics assessed for sex/gender differences)	Percent of brains/individuals with substantial variability: both male-end and female-end features	Percent of brains/individuals with internal consistency: only female-end (F), only intermediate (I), or only male-end (M) (all, ♂, ♀)	Average (SD) percent of features at the female-end (F) and male-end (M) zones (♂, ♀) and Cohen's *d* of the sex/gender difference
First sample, VBM	♂: 18–79, 31.5 (12.0) ♀: 18–75 28.9 (10.4)	10 (116) $0.70 < \|d\| \leq 0.84$ all $P < 0.0001$	All: 35% ♂: 35% ♀: 34%	F: 0.4, 0.0, 0.6% I: 3.6, 3.6, 3.6% M: 2.0, 5.0, 0%	F: 13 (17), 33 (25) $d = 0.95$* M: 33 (30), 10 (15) $d = -0.96$*
1000,† VBM	♂: 18–74, 28.8 (14.3) ♀: 18–78, 26.8 (10.4)	10 (116) $0.51 < \|d\| \leq 0.69$ all $P < 0.0001$	All: 39% ♂: 37% ♀: 40%	F: 0.1, 0.0, 0.2% I: 2.3, 3.3, 1.6% M: 2.9, 5.3, 1.2%	F: 15 (19), 33 (25) $d = 0.83$* M: 33 (31), 17 (23) $d = -0.61$*
1000,† VBM 18–26 subsample	♂: 18–26, 21.5 (1.9) ♀: 18–26, 21.5 (2)	9‡ (116) $0.46 < \|d\| \leq 0.60$ all $P < 0.0001$	All: 53% ♂: 47% ♀: 55%	F: 0.3, 0.4, 0.2% I: 1.3, 0.8, 1.6% M: 0.8, 1.2, 0.5%	F: 16 (19), 33 (24) $d = 0.79$* M: 33 (25), 19 (20) $d = -0.62$*
NKI, SBA, cortical thickness	♂: 13–83, 41.0 (20.3) ♀: 12–85, 48.7 (17.4)	7 (68) $0.41 < \|d\| \leq 0.56$ all $P < 0.002$	All: 24% ♂: 21% ♀: 26%	F: 4.5, 2.0, 5.9% I: 2.2, 2.0, 2.4% M: 3.7, 8.0, 1.2%	F: 21 (27), 33 (29) $d = 0.42$* M: 33 (34), 15 (22) $d = -0.64$*
NKI, SBA, volume		12 (168) $0.94 < \|d\| \leq 1.04$ all $P < 0.0001$	All: 23% ♂: 25% ♀: 21%	F: 1.5, 0.0, 2.3% I: 3.3, 2.9, 3.6% M: 0.7, 1.9, 0.0%	F: 9 (16), 33 (27) $d = 1.05$* M: 33 (27), 16 (15) $d = -1.13$*
DTI fractional anisotropy	♂: 17–43, 24.8 (4.6) ♀: 18–57, 26.3 (7.0)	11 (116) $0.73 < \|d\| \leq 1.05$ all $P < 0.0001$	All: 25% ♂: 29% ♀: 20%	F: 2.2, 0.0, 4.3% I: 2.9, 2.9, 2.9% M: 0.7, 1.4, 0.0%	F: 9 (15), 33 (33) $d = 0.93$* M: 33 (29), 12 (20) $d = -0.83$*
DTI connectivity		7 (4,005) $0.66 < \|d\| \leq 0.96$ all $P < 0.00017$§	All: 48% ♂: 52% ♀: 43%	F: 0.0, 0.0, 0.0% I: 0.7, 0.0, 1.4% M: 0.0, 0.0, 0.0%	F: 14 (16), 33 (18) $d = 1.15$* M: 33 (20), 9 (11) $d = -1.53$*
MADICS	♂: 20–23, 21.6 (0.7) ♀: 20–23, 21.3 (0.6)	7 (31) $0.43 < \|d\| \leq 0.77$ all $P < 0.0001$	All: 59% ♂: 64% ♀: 56%	F: 0.0, 0.0, 0.0% I: 1.8, 1.1, 2.1% M: 0.0, 0.0, 0.0%	F: 17 (15), 32 (18) $d = 0.92$* M: 32 (17), 13 (14) $d = -1.23$*
ADD Health	♂: 18–28, 22.4 (1.9) ♀: 18–28, 22.1 (1.9)	8 (26) $0.41 < \|d\| \leq 0.57$ all $P < 0.0001$	All: 70% ♂: 81% ♀: 62%	F: 0.0, 0.0, 0.0% I: 0.1, 0.2, 0.03% M: 0.0, 0.0, 0.0%	F: 27 (16), 45¶ (19) $d = 1.04$* M: 29¶ (17), 13 (13) $d = -1.01$*
Carothers & Reis' data	21.15 (7.68)	10 (10) $1.0 < \|d\| \leq 2.02$ all $P < 0.0001$	All: 55% ♂: 65% ♀: 48%	F: 0.4, 0.0, 0.6% I: 0.4, 0.9, 0.0% M: 0.4, 0.9, 0.0%	F: 11 (11), 48¶ (20) $d = 2.27$* M: 41¶ (17), 8 (10) $d = -2.42$*

♂, males; ♀, females.
*Statistically significant difference, $P < 0.0001$.
†1000 = 1000 Functional Connectomes Project.
‡If we were to include only regions with |Cohen's d| > 0. 51 as in the analysis of the entire sample, only three regions would have been included.
§Of the seven connections included, two were statistically significant and five approached significance.
¶Note the deviation from 33%, which reflects the inability to define the cutoff at 33% for some of the variables included in the analysis (see *Methods* for details).

Results

T1-weighted images of 169 females and 112 males (Table 1) were preprocessed and assessed for gray matter volume using voxel-based morphometry (VBM) (Fig. 1 *A* and *B*). Of the 116 regions of gray matter defined using the Automated Anatomical Labeling atlas (AAL) (11) (Fig. 1*C*), 10 regions showing the largest sex/gender differences (|Cohen's d| > 0.70, the largest |d| was 0.84; all $P < 0.0001$) were included in subsequent analyses (Table S1). Using the actual distributions of males and females in the sample, "male-end" and "female-end" zones were arbitrarily defined as the scores of the 33% most extreme males and females, respectively, and an "intermediate" zone was defined as the area in-between these two (Fig. 1*D*, we use the terms "male-end"/"female-end" as a shorthand for "the end of the continuum in which males/females are more prevalent," respectively; note that our method of categorizing the continuum into three discrete classes inherently places some females at the "male-end" and some males at the "female-end"). Fig. 1*E* presents the gray matter volume of the 10 regions in each of the females (*Left*) and in each of the males (*Right*) using a color volume scale (as shown in Fig. 1*D*), and clearly illustrates the lack of internal consistency in most brains. The latter is also evident in Fig. 1*F*, which presents the number of "female-end" (x axis) and "male-end" (y axis) characteristics in females (red) and males (green). The circles at the (10,0), (0,10), and (0,0) coordinates represent individuals with only "female-end", only "male-end", or only "intermediate" characteristics, respectively; All other circles on the x and y axes represent individuals who have either "female-end" or "male-end" characteristics, as well as "intermediate" characteristics; The rest of the circles represent individuals with substantial variability, having both regions at the "male-end" and regions at the "female-end". Thirty-five percent of brains showed substantial variability, and only 6% of brains were internally consistent (see Table 1 for more details). Notably, additional definitions of the "male-end" and "female-end" zones (50%, 20%, and 10%) similarly revealed a much higher prevalence of brains showing substantial variability compared with brains showing internal consistency (Table S2). Importantly, substantial variability is not a result of the overlap between females and males in each of the brain regions, as evident in Fig. S1*A*, which depicts the results that would have been obtained for these data under perfect internal consistency (for comparison, Fig. S1 *B*–*D* depicts these data under internal consistency with different degrees of random noise, and Fig. S1*E* under no internal consistency).

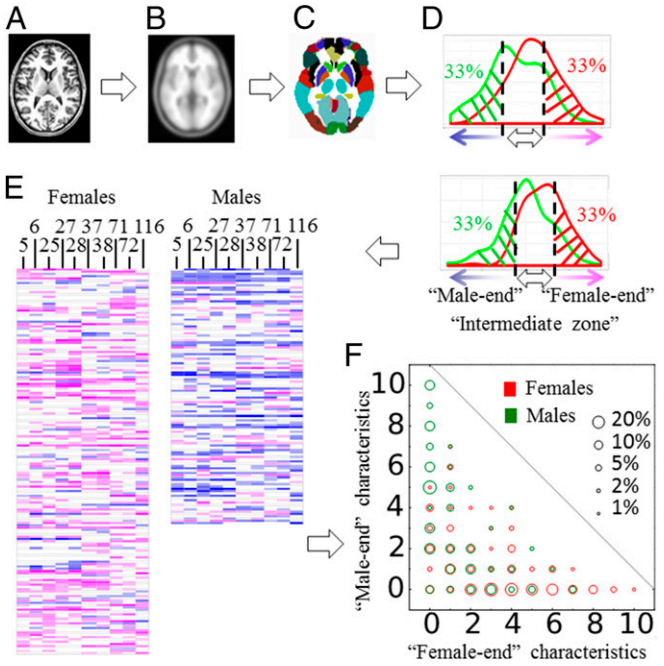

Fig. 1. Assessing internal consistency in the human brain. (*A*) T1-weighted images were normalized and segmented using (*B*) the MNI template. (*C*) Voxels were mapped into 116 regions according to the AAL atlas. (*D*) The distribution of the gray matter volume in females (red) and males (green) of two of the region showing the largest sex/gender differences [left hippocampus (*Upper*) and left caudate (*Lower*)]. A continuous color representation of the degree of "maleness" and "femaleness" was created separately for each of the 10 regions showing the largest sex/gender difference. Volumes falling in the "intermediate" zone are colored in white; volumes in the "male-end" and in the "female-end" zones are colored using continuous blue-white and pink-white scales, respectively. (*E*) The degree of "maleness-femaleness" of each region for each female (*Left*) and male (*Right*). Each horizontal line represents the brain of one subject and each column represents a single brain region. The number above each column corresponds to the region's number in the AAL atlas and in Table S1. (*F*) A bivariate scattergram (created using the Matplotlib library of Python) of the number of regions at the "female-end" (x axis) and at the "male-end" (y axis) in females (red) and males (green). The number of regions at the "intermediate" zone is not depicted because the number of "male-end", "intermediate", and "female-end" features always adds up to the number of features included in the analysis (which is the highest value on the x and y axes of the graph). The size of each circle is proportional to the percent of individuals from the same sex/gender category with an identical score on the two measures.

A similar pattern of results was obtained when we repeated the same analysis on the data of 495 females and 360 males obtained from the 1000 Functional Connectomes Project (12) (Table 1, Table S2, and Fig. 2*A*). It is noteworthy that the two datasets revealed a similar pattern of results even though, of the 10 regions included in each analysis, only two were common (Table S1). To exclude the possibility that this pattern of results was a result of the large age range included in the two samples, as sex/gender differences have been reported to differ in different age groups (13–15), we repeated the same analysis on 625 individuals from the 1000 Functional Connectomes sample (385 females, 240 males) 18–26 y of age. Substantial variability was seen in 52% of brains, whereas internal consistency was evident in only 2.4% (Table 1, Tables S1 and S2, and Fig. 2*B*).

To test whether the pattern of results obtained in the two datasets was dependent on the type of analysis of the imaging data (VBM), and in particular on the "correction" for brain size included in the latter procedure, we performed the same analysis on a subgroup of the Nathan Kline Institute (NKI) enhanced sample (167 females, 100 males), whose T1-weighted images were preprocessed and manually corrected for cortical surface-based analysis. Using the FreeSurfer software package (Athinoula A. Martinos Center for Biomedical Imaging, Harvard University), 68 cortical regions were delineated in the native surface space (16), and the mean cortical thickness values (gray-white matter boundary to pial boundary) were calculated. This delineation method allows for direct comparison of regional and whole-brain cortical features in contrast to the VBM analysis described above, which includes normalization to a volumetric template space and "correction" for brain size. Also here, substantial variability was more common than internal consistency (24% and 10.5%, respectively; Table 1, Tables S2 and S3, and Fig. 2*C*). An analysis of the noncorrected volume of these 68 cortical regions, 23 subcortical regions of gray matter, and 77 regions of white matter revealed that substantial variability was much more common than internal consistency (23% and 2.2%, respectively; Table 1, Tables S2 and S3, and Fig. 2*D*). This latter finding is of particular interest because the sex/gender differences observed in this dataset were particularly large.

To test whether the pattern of results obtained was dependent on the type of imaging (T1-weighted images), we performed the same analysis on diffusion tensor imaging (DTI) indices of 69 females and 69 males. Using the AAL atlas used for the VBM, we calculated the average fractional anisotropy and mean diffusivity

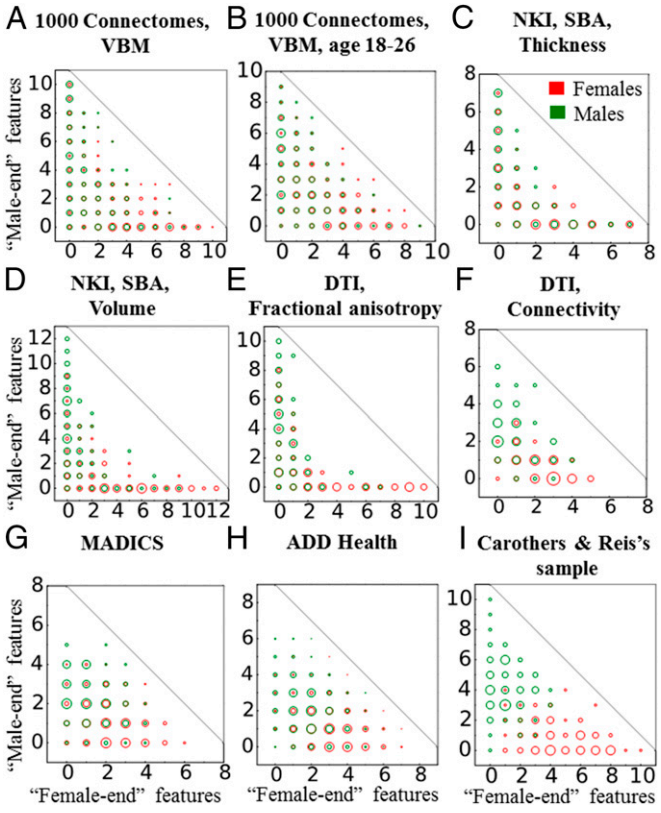

Fig. 2. Distributions of "female-end" and "male-end" features in females and males. Bivariate scattergrams of the number of features at the "female-end" (*x* axis) and at the "male-end" (*y* axis) in females (red) and males (green) in the 1000 Functional Connectomes Project sample, VBM analysis (*A*), subsample of the 18–26 y olds from the 1000 Functional Connectomes Project sample, VBM analysis (*B*), the NKI, SBA, cortical thickness (*C*), and volume of cortical, subcortical, and white matter regions (*D*), the University of Zurich DTI data, fractional anisotropy (*E*) and connectivity (*F*), MADICS sample (*G*), ADD Health sample (*H*), and Carothers and Reis's sample (*I*). The size of each circle is proportional to the percent of individuals from the same sex/gender category with an identical score on the two measures.

for each of the 116 regions of gray matter. The sex/gender differences in mean diffusivity were too small to survive the correction for multiple comparisons. Analysis of the fractional anisotropy data revealed substantial variability in 28% of brains and internal consistency in 5.8% (Table 1, Tables S1 and S2, and Fig. 2*E*).

We then used the DTI data to assess brain connectivity. Using deterministic fiber tractography, we estimated the connectivity strength between 90 regions of gray matter defined with the AAL atlas. Analysis of the 7 connections (of 4,005 connections) with the largest sex/gender differences revealed substantial variability in 48% of brains and internal consistency in only 0.7% (all of which were in the "intermediate" zone; Table 1, Tables S2 and S4, and Fig. 2*F*).

The low degree of internal consistency observed here in the human brain agrees well with studies demonstrating that humans often possess both "masculine" and "feminine" psychological characteristics (that is, personality traits, attitudes, interests, and behaviors that show sex/gender differences). Early attempts in the first half of the 20th century to measure masculinity-femininity using specially constructed scales have already revealed low or absent correlations between subscales measuring different characteristics of gender (17). Similar findings have led Janet Spence (18) to conclude that humans possess an array of masculine and feminine traits that cannot be captured using a uni-dimensional (masculinity-femininity) or a bidimensional (masculinity × femininity) model. However, to date, only a few studies have followed this line of research (19–21), and those that have, only assessed a small number of variables. We therefore used the same approach described above to analyze two open datasets that provide data on many psychological variables for a large number of subjects: the Maryland Adolescent Development In Context Study (MADICS) (22) and the National Longitudinal study of Adolescent Health (ADD Health) (23).

Of the different measures of behavior, personality characteristics, and attitudes available in MADICS, we analyzed data of seven variables with the largest sex/gender differences (Table S5) of 382 females and 188 males (Table 1). Substantial variability was evident in 59% of subjects and internal consistency in only 1.8% (all of which were in the "intermediate" zone; Fig. 2*G* and Fig. S2). Very similar results were obtained when analyzing the data of 2,239 males and 2,621 females from the ADD Health study, which is a study of a US-representative sample of adolescents. Substantial variability was evident in 70% of subjects and internal consistency in only 0.1% (all of which were in the "intermediate" zone; Fig. 2*H*, Table 1, and Table S5).

Last, we contrasted our analysis of the different brain- and gender-related datasets with a similar analysis of one of Carothers and Reis' (22) behavioral datasets, which was unique in that it was the only behavioral dataset in which humans could be meaningfully grouped into two distinct categories on the basis of their sex (19). The dataset included 10 highly gender-stereotyped activities (e.g., playing videogames and watching talk shows; Table S5) specifically selected to differentiate between females (n = 157) and males (n = 106) of this subculture (introductory-level psychology class students at a large Midwestern American university). Accordingly, the sex/gender differences were very large ($1.00 < |\text{Cohen's } d| < 2.02$), and the distribution of several of the variables was bimodal and highly skewed at both ends. This dataset was also unique in our analysis (compare Fig. 2*I* to Fig. 2 *A*–*H*), in that there was almost no overlap between females and males in the possible combinations of the number of "female-end" and "male-end" characteristics, as in most subjects the number of "consistent" characteristics was higher than the number of "nonconsistent" characteristics. Interestingly, also in this sample, 55% of subjects showed substantial variability and only 1.2% were internally consistent (Table 1, Fig. 2*I*, and Fig. S2). In other words, even when considering highly stereotypical gender behaviors, there are very few individuals who are consistently at the "female-end" or at the "male-end", but there are many individuals who have both "female-end" and "male-end" characteristics. Furthermore, although one's sex is enough to predict whether this person would have more "female-end" or more "male-end" characteristics, it is not enough to predict this person's specific combination of "female-end" and "male-end" characteristics (Fig. S2) (for further discussion of the question of prediction, see ref. 4).

Discussion

Consistent with previous findings (14, 15), our analysis of the structure of the human brain, which included most regions of gray and white matter, as well as measures of connectivity, revealed many nondimorphic group-level sex/gender differences in brain structure. There was extensive overlap of the distributions of females and males for all brain regions and connections assessed, irrespective of the type of sample, measure, or analysis (including analysis of absolute brain volumes). This extensive overlap undermines any attempt to distinguish between a "male" and a "female" form for specific brain features. Rather, the forms that are evident in most females are also the ones evident in most males (Fig. 1*D*). It is therefore more appropriate and informative to refer to measures of the brain in quantitative ways (Fig. 3) rather than in qualitative ways (e.g., "male", "female" form). Another noteworthy observation is that the size of the sex/gender

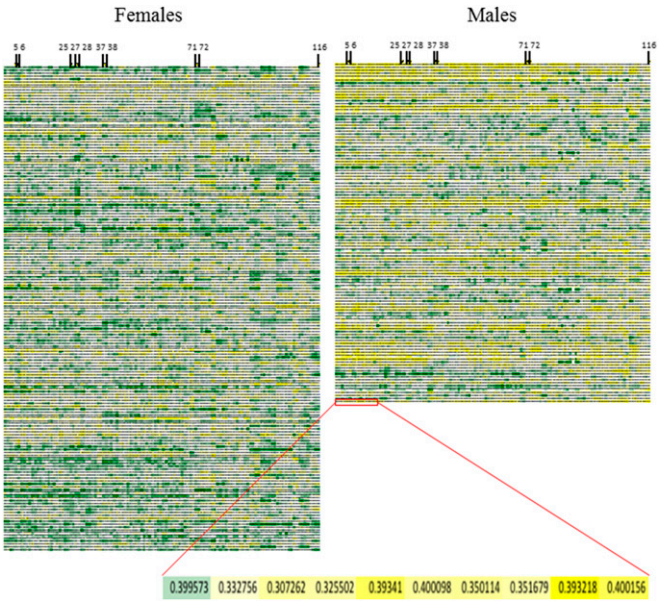

Fig. 3. The human brain mosaic. The gray matter volume of all 116 regions of gray matter in females (*Left*) and in males (*Right*) from the first sample is represented using a continuous high-low (green-white-yellow) scale. Each horizontal line represents the brain of a single subject and each column represents a single brain region. The continuous high-low scale represents the relative volume of a brain region in a given brain relative to the volume of this brain region in all other brains (i.e., within a column). The regions that showed the largest sex/gender differences and were included in the internal consistency analysis are marked with a black bar. The number above each bar corresponds to the region's number in the AAL atlas and in Table S1. (*Inset*) Magnification of a small part of a horizontal line (i.e., a single brain). The number in each colored cell is the volume of this region for this brain.

difference in some regions varied considerably between different datasets (Table S1). This finding is in line with previous reports that the existence and direction of sex/gender differences may depend on environmental events and developmental stage (4, 5).

The novel aspect of the present study is the addition of another level of consideration to current thinking about the relation between sex and the brain. Specifically, this study is the first, to our knowledge, to move beyond the level of sex/gender differences in single brain elements (e.g., the volume of a brain region) to the level of the brain as a whole, by assessing internal consistency in the degree of "maleness-femaleness" of different elements within a single brain. Our results demonstrate that even when analyses are restricted to a small number of brain regions (or connections) showing the largest sex/gender differences, internal consistency is rare and is much less common than substantial variability (i.e., being at the one end of the "maleness-femaleness" continuum on some elements and at the other end on other elements). This finding was independent of sample, age, type of imaging, method of analysis of the imaging data, and the specific definition of the end of the continuum (i.e., the percent of individuals included in the "male-end" and "female-end" zones; Table S2).

Our conclusion that substantial variability is much more common than internal consistency in the human brain may have implications for current theories of the sexual differentiation of the brain and, in particular, for the classic view that the female brain is the default pathway and the male brain is a differentiation away from that default (9). On this view, one could expect that there may be greater variability in males compared with females in the degree of differentiation, leading to a higher prevalence of substantial variability and of "nonconsistent" characteristics in males compared with females. Our data, however, do not support this view as the proportion of males and of females with substantial variability was not statistically different in any of the seven datasets, and the average proportion of "nonconsistent" characteristics was significantly higher in males compared with females in only one of the seven datasets (connectivity, $P = 0.035$). Thus, our findings that substantial variability is much more prevalent than internal consistency together with the lack of evidence for consistent sex/gender differences in the propensity to exhibit substantial variability do not support the classic view. Our findings are in line, however, with more recent thinking that masculinization and feminization are two independent processes and that sexual differentiation progresses independently in different brain tissues, "enabling genetically and environmentally induced variation in sexual differentiation of different tissues within a single brain" (4, p. 4; 10).

Our study demonstrates that although there are sex/gender differences in brain structure, brains do not fall into two classes, one typical of males and the other typical of females, nor are they aligned along a "male brain–female brain" continuum. Rather, even when considering only the small group of brain features that show the largest sex/gender differences, each brain is a unique mosaic of features, some of which may be more common in females compared with males, others may be more common in males compared with females, and still others may be common in both females and males. The heterogeneity of the human brain and the huge overlap between the forms that brains of males and brains of females can take can be fully appreciated when looking at the entire brain (Fig. 3 and Figs. S3 and S4).

In accordance with the brain data, our analyses of gender-related data revealed extensive overlap between females and males in personality traits, attitudes, interests, and behaviors. Moreover, we found that substantial variability of gender characteristics is highly prevalent, whereas internal consistency is extremely rare, even for highly gender-stereotyped activities (Carothers and Reis' data). These findings are in line with previous reports that sex/gender differences in abilities and qualities are mostly nonexistent or small, that there is extensive overlap between the distribution of males and females also in behaviors, interests, occupation preferences, and attitudes that show larger sex/gender differences (24, 25), and that there are no or only weak correlations between gender characteristics (18, 20, 21). Thus, most humans possess a mosaic of personality traits, attitudes, interests, and behaviors, some more common in males compared with females, others more common in females compared with males, and still others common in both females and males.

Conclusions

The lack of internal consistency in human brain and gender characteristics undermines the dimorphic view of human brain and behavior and calls for a shift in our conceptualization of the relations between sex and the brain. Specifically, we should shift from thinking of brains as falling into two classes, one typical of males and the other typical of females, to appreciating the variability of the human brain mosaic. Scientifically, this paradigm shift entails replacing the currently dominant practice of looking for and listing sex/gender differences with analysis methods that take into account the huge variability in the human brain (rather than treat it as noise), as well as individual differences in the specific composition of the brain mosaic. At the social level, adopting a view that acknowledges human variability and diversity has important implications for social debates on long-standing issues such as the desirability of single-sex education and the meaning of sex/gender as a social category.

Methods

Data Collection and Preparation for Analysis.

Brain-related data. Data were obtained from four sources: Tel-Aviv University (the first brain-related dataset), University of Zurich (26) (DTI data), the 1000 Functional Connectomes Project (12), and the NKI enhanced sample

(FreeSurfer analysis). For details of the imaging protocols, the datasets included from the 1000 Functional Connectomes Project, and the analysis of the images, see *SI Methods*.

Gender-related data. Data were obtained from the MADICS (22), ADD Health (27), and from Harry Reis (the Carothers and Reis's sample). We used data from the sixth wave of MADICS and the third wave of ADD Health because these waves included data of young adults (between 20 and 23 y old in MADICS and between 18 and 28 y old in ADD Health) on many variables that are known to show sex/gender differences, such as personality traits, relationships, activities, and attitudes. For further details, see *SI Methods*.

Data Analysis. For each dataset, we calculated the significance [using the false discovery rate (FDR) method to correct for multiple comparisons] (28) and the effect size [Cohen's d = (Mfemales − Mmales)/$\sqrt{([(SDfemales^2 + SDmales^2)]/2)}$] of the sex/gender difference for every variable. In calculating Cohen's d, we weighted the variances according to the proportion of males and females in the population (~50%) and not according to the actual proportion in each dataset so as not to bias the estimate of the size of the difference due to the nonequal number of males and females in most of our datasets. In each dataset, of the variables showing significant sex/gender differences, subsequent analyses used only the variables showing the largest sex/gender differences, because in large datasets, as were some of the datasets we used, even very small differences with a great overlap between females and males are significant.

For each of the variables chosen for further analysis, we defined "male-end" and "female-end" zones as the scores of the 33% most extreme males and females, respectively, and an "intermediate" zone in between these two (Fig. 1*D*). For gender-related variables with discrete scoring, we chose as the "male-end"/"female-end" zone the zone that was nearest to 33%. Note that for such variables, the proportion of males at the "male-end" may not equal the proportion of females at the "female-end". Once the three zones were defined for each variable, we defined for each subject his/her form in each of the variables and then defined for each subject whether s/he was internally consistent at the "male-end," "female-end," or "intermediate" zone or whether s/he had substantial variability (having at least one characteristic at the "female-end" and one characteristic at the "male-end"). In addition, we calculated each subject's proportion of "female-end" and "male-end" characteristics. The Student t test was used to compare the mean proportion of "nonconsistent" characteristics in females and males, and the two-proportion z-test was used to compare the proportion of males and females showing substantial variability.

Creating a Continuous Color Code.

Pink-white-blue. The scale was created separately for each brain region (and for each gender characteristic) on the basis of the definitions of its "male-end", "female-end", and "intermediate" zones. Values in the "female-end" ("male-end") zone were colored using the three-color scale conditional formatting function in Excel (Version 14.5.2), with the most extreme score defined as pink (blue), and the score bordering the "intermediate" zone defined as white (Fig. 1*D*).

Green-white-yellow. The scale was created separately for each brain region using the three-color scale conditional formatting function in Excel. The highest score was defined as green, the lowest score was defined as yellow, and the middle score was defined as white. In samples with equal numbers of females and males, the middle score was the median. In the other samples, the middle score was chosen so that the proportion of males on one side of this score equals the proportion of females on the other side to not bias the estimate of the middle of the distribution due to the nonequal number of males and females.

ACKNOWLEDGMENTS. We thank Dr. Bobbi Carothers (Washington University in St. Louis) and Prof. Harry Reis (University of Rochester) for allowing us to use their data and Prof. Reis for stimulating discussions of the mosaic hypothesis. This research used the Maryland Adolescent Development In Context Study of Adolescent Development in Multiple Contexts, 1991–1998 (Log 1066) dataset (made accessible in 2000, numeric data files). These data were collected by Jacqueline S. Eccles (Producer) and are available through the Henry A. Murray Research Archive of the Institute for Quantitative Social Science at Harvard University (Distributor). Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. This research used data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill and funded by Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant P01-HD31921, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the National Longitudinal study of Adolescent Health (Add Health) website (www.cpc.unc.edu/addhealth). No direct support was received from Grant P01-HD31921 for this analysis. This work was supported by Swiss National Science Foundation Grants 320030-120661, 320030B-138668, 20030B-138668, and 4-62341-05 and European Union Future and Emerging Technologies Integrated Project Presence: Research Encompassing Sensory Enhancement, Neuroscience, Cerebral-Computer Interfaces and Application (PRESENCCIA) Grant 27731 (to L.J.).

1. Richardson SS (2013) *Sex Itself: The Search for Male and Female in the Human Genome* (Univ of Chicago Press, Chicago).
2. Schiebinger L (1989) *The Mind Has No Sex?: Women in the Origins of Modern Science* (Harvard Univ Press, Cambridge, MA).
3. Ingalhalikar M, et al. (2014) Sex differences in the structural connectome of the human brain. *Proc Natl Acad Sci USA* 111(2):823–828.
4. Joel D (2011) Male or female? Brains are intersex. *Front Integr Neurosci* 5:57.
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 Differ* 3(1):27.
6. Eliot L (2011) Single-sex education and the brain. *Sex Roles* 69(7):363–381.
7. Jordan-Young R, Rumiati RI (2012) Hardwired for sexism? Approaches to sex/gender in neuroscience. *Neuroethics* 5(3):305–315.
8. Maney DL (2014) *Just Like a Circus: The Public Consumption of Sex Differences. Current Topics in Behavioral Neurosciences* (Springer-Verlag, Berlin).
9. Arnold AP (2009) The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Horm Behav* 55(5): 570–578.
10. McCarthy MM, Arnold AP (2011) Reframing sexual differentiation of the brain. *Nat Neurosci* 14(6):677–683.
11. Tzourio-Mazoyer N, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15(1):273–289.
12. Biswal BB, et al. (2010) Toward discovery science of human brain function. *Proc Natl Acad Sci USA* 107(10):4734–4739.
13. Garcia-Falgueras A, Ligtenberg L, Kruijver FP, Swaab DF (2011) Galanin neurons in the intermediate nucleus (InM) of the human hypothalamus in relation to sex, age, and gender identity. *J Comp Neurol* 519(13):3061–3084.
14. Lenroot RK, Giedd JN (2010) Sex differences in the adolescent brain. *Brain Cogn* 72(1): 46–55.
15. McCarthy MM, Konkle AT (2005) When is a sex difference not a sex difference? *Front Neuroendocrinol* 26(2):85–102.
16. Fischl B, et al. (2004) Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14(1):11–22.
17. Lippa RA (2005) *Gender, Nature, and Nurture* (Psychology Press, New York), 2nd Ed.
18. Spence JT (1993) Gender-related traits and gender ideology: Evidence for a multifactorial theory. *J Pers Soc Psychol* 64(4):624–635.
19. Carothers BJ, Reis HT (2013) Men and women are from Earth: Examining the latent structure of gender. *J Pers Soc Psychol* 104(2):385–407.
20. Egan SK, Perry DG (2001) Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Dev Psychol* 37(4):451–463.
21. Koestner R, Aube J (1995) A multifactorial approach to the study of gender characteristics. *J Pers* 63(3):681–710.
22. Eccles JS (1999) *MADICS Study of Adolescent Development in Multiple Contexts, 1991-1998* (Murray Research Archive, Cambridge, MA).
23. Harris KM (2009) *The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009* (Carolina Population Center, Univ of North Carolina at Chapel Hill, Chapel Hill, NC).
24. Hyde JS (2005) The gender similarities hypothesis. *Am Psychol* 60(6):581–592.
25. Hyde JS (2014) Gender similarities and differences. *Annu Rev Psychol* 65:373–398.
26. Hänggi J, Fövenyi 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.
27. Population Health UNC (2003) National Longitudinal Study of Adolescent Youth. Available at www.cpc.unc.edu/projects/addhealth. Accessed December 17, 2013.
28. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc, B* 57(1):289–300.
29. Good CD, et al. (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* 14(1 Pt 1):21–36.
30. Smith SM, et al. (2004) Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23(Suppl 1):S208–S219.
31. Smith SM, et al. (2006) Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31(4):1487–1505.
32. Behrens TEJ, et al. (2003) Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn Reson Med* 50(5):1077–1088.
33. Zalesky A, et al. (2010) Whole-brain anatomical networks: Does the choice of nodes matter? *Neuroimage* 50(3):970–983.

Supporting Information

Joel et al. 10.1073/pnas.1509654112

SI Methods

Brain Imaging. Data from Tel-Aviv were obtained with a 3-T (GE) MRI system. T1-weighted images were acquired with a 3D spoiled gradient-recalled echo sequence with the following parameters: up to 160 axial slices (whole brain coverage), TR/TE = 9/3 ms, resolution, $1 \times 1 \times 1$ mm^3, and scan time 4 min. Data from Zurich were obtained with a 3-T Philips Achieva MRI scanner (Philips Medical Systems). One diffusion-weighted spin-echo echo-planar imaging sequence was applied with the following parameters: 75 axial slices (whole brain coverage), TR/TE = 13.5 s/55 ms, spatial resolution $2 \times 2 \times 2$ mm^3, b-value = 1,000 s/mm^2, 32 noncollinear diffusion directions, one non–diffusion-weighted image, and scan time 10 min.

Datasets from the 1000 Functional Connectomes Project were included from the following sites: Atlanta; Baltimore; Beijing; Berlin; Cambridge, MA; the International Consortium for Brain Mapping (ICBM), Montreal; Leiden, The Netherlands; Milwaukee; Munich; Newark, NJ; New York City; Orangeburg, NY; Oulu, Finland; Palo Alto, CA; Queensland, Australia; St. Louis). Other datasets were excluded from analysis due to preprocessing difficulties or because they included skull-stripped images, as we wanted to avoid systematic differences between datasets.

Analysis of T1-weighted images.

Volume-based analysis. Images were analyzed using MATLAB (MathWorks) and SPM8 (Wellcome Department of Cognitive Neurology; www.fil.ion.ucl.ac.uk/spm). Gray matter volume was assessed with the optimized voxel-based morphometry (VBM) protocol (29), using the standard segmentation and registration tools available in the software. Images were normalized, segmented, modulated, and smoothed with a 12-mm Gaussian kernel. Voxels were mapped into 116 regions according to the AAL atlas (11), and mean gray matter volume was calculated for each region for each participant. For the full list of regions assessed see neuro.imm.dtu.dk/wiki/Automated_Anatomical_Labeling.

Surface-based analysis. The FreeSurfer software package (Athinoula A. Martinos Center for Biomedical Imaging, Harvard University; surfer.nmr.mgh.harvard.edu/fswiki) was used to generate the surface representations of the cortex and to delineate 68 regions (regions 1001–1003, 1005–1035, 2001–2003, and 2005–2035 in the list of all regions that can be delineated using FreeSurfer, which appears in www.slicer.org/slicerWiki/index.php/Documentation/4.1/SlicerApplication/LookupTables/Freesurfer_labels). For each participant, we calculated the average cortical thickness and total cortical volume for each of these regions, as well as the volumes of 77 white matter regions (regions 7, 46, 251–255, 3001–3003, 3005–3035, 4001–4003, 4005–4035, and 5001–5002) and of 23 subcortical structures (4–5, 8, 10–18, 26, 28, 47, 49–54, 58, and 60).

Diffusion tensor imaging analysis. Data analysis was performed with FSL (30) (www.fmrib.ox.ac.uk/fsl) using FMRIB's diffusion toolbox (32) and following the standard preprocessing steps implemented in tract-based spatial statistics (31) with default parameters, but excluding the skeletonization step. Voxels were mapped into 116 regions according to the AAL atlas (11), and average fractional anisotropy and average mean diffusivity were calculated for each region and for each participant.

For the analysis of brain connectivity, we preprocessed the diffusion-weighted connectivity data with FDT (32). For deterministic fiber tractography, we used the Diffusion Toolkit (DTK; trackvis.org/) and TrackVis software (trackvis.org/). The connectivity matrix, based on the number of reconstructed streamlines between 90 AAL regions of interest (the 26 cerebellar AAL regions were excluded because the cerebellum was not covered completely in some subjects), was computed using MATLAB scripts written by Andrew Zalesky (33). A more detailed description of the connectivity methods applied in the present study can be found elsewhere (26).

Gender-Related Data. Datasets and questionnaires that had specific instructions for the creation of variables were analyzed as instructed. Additional variables were created by grouping together variables intended to measure a single construct (e.g., depression) or that content-wise seemed to assess related constructs (e.g., different measures of the amount of house work) and showed a high correlation (in case there were only two such variables) or Cronbach's α (in case there were multiple items). This approach was taken to reduce intervariable correlations that reflect measuring the same construct in different ways, rather than the consistent effects of sex/gender on different variables within an individual. Because there were many missing data, we first chose the variables showing the largest sex/gender differences over the entire sample and then performed the analysis only on subjects who had no more than one (MADICS) or two (ADD Health) missing data for these variables. The data reported in Table 1 and Table S5 were calculated over the subjects included in the final analysis.

Hypothetical Distributions of "Male-End" and "Female-End" Scores Under Different Degrees of Internal Consistency. These simulations were created on the basis of the data of the first sample. For all simulations, we created 10,000 "male brains" and 10,000 "female brains", each with 10 "regions" (one for each brain region in the actual data). The score for each "brain" on a "region" was a number between 1 and 10,000 (because in our method of determining the "male-end" and "female-end" only the order of scores within a region matters). Next, for the "female brains", for each "region", we determined the "female-end" zone as the lowest 33% scores and the "male-end" zone as the highest X% scores, with X for each "region" taken from the actual percent of females with a "male-end" region for the simulated brain region in the actual data (because the degree of overlap between females and males in the 10 brain regions in the actual data were not identical, the percent of females in the "male-end" differed between the 10 regions ranging between 0.14 and 0.19). The same was done for the "male brains" (the percent of males in the "female-end" in the 10 regions ranged between 0.12 and 0.20). Last, we determined for each "brain" the number of "male-end", "intermediate", and "female-end" "regions", and plotted a bivariate scattergram of the number of "regions" at the "female-end" (x axis) and at the "male-end" (y axis) in "females" (red) and "males" (green, as was done in Fig. 1*F*). Perfect internal consistency was created by giving the same score (i.e., the same location along the "femaleness-maleness" continuum) for all "regions" of a single "brain" (e.g., if a "brain" had a score of 150 in the first "region", it also had a score of 150 on the other nine "regions"). Thus, the correlations between all possible pairs of two "brain regions" were 1. Internal consistency with noise was simulated by adding noise to the perfectly consistent "brains". Noise was added by randomly adding to each score a number between −2,500 and +2,500 (average absolute noise was 1,250, average correlation between all possible pairs of two "brain regions" was 0.80), or between −3,333 and +3,333 (average absolute noise was 1,666.5, average correlation between all possible pairs of two "brain regions" was 0.70), or between −5,000 and

+5,000 (average absolute noise was 2,500, average correlation between all possible pairs of two "brain regions" was 0.50). No internal consistency was simulated by randomly assigning a number between 1 and 10,000 for each of the 10 "regions" in each of the 10,000 "female brains" and each of the 10,000 "male brains". The average correlation between all possible pairs of two "brain regions" was 0.0009.

SI Results

Hypothetical Distributions of "Male-End" and "Female-End" Scores Under Different Degrees of Internal Consistency. Under absolutely no internal consistency, ~80% of "brains" showed substantial variability compared with 0.1% that showed internal consistency (Fig. S1*E*). Under perfect internal consistency, ~93% of "brains" showed internal consistency compared with 0% that showed substantial variability (Fig. S1*A*; the percent of internally consistent "brains" is not 100% because the percent of subjects in the "nonconsistent"-end differs between the 10 regions, so that even under perfect consistency there are some individuals who have a combination of "nonconsistent"-end and "intermediate" regions). Adding noise disrupted internal consistency more than it increased substantial variability. Thus, moving from perfect internal consistency to higher degrees of noise decreased internal consistency from ~93% (Fig. S1*A*) to ~24% (±2,500; Fig. S1*B*), ~9% (±3,333; Fig. S1*C*), and ~1.4% (±5,000; Fig. S1*D*) of "brains", while increasing substantial variability from 0% (Fig. S1*A*) to 0.005% (Fig. S1*B*), ~3.5% (Fig. S1*C*), and ~20.5% (Fig. S1*D*) of "brains". A comparison of the ±5,000 simulated condition to the actual data, in which 6% of brains showed internal consistency and 35% showed substantial variability, suggests that noise cannot explain the pattern of results we obtained, because less noise is expected to account for the percent of internal consistency but more noise to account for the percent of substantial variability. Future studies will explore the type of model that may account for the pattern of results obtained in the first sample, as well as in the other datasets analyzed in the present study.

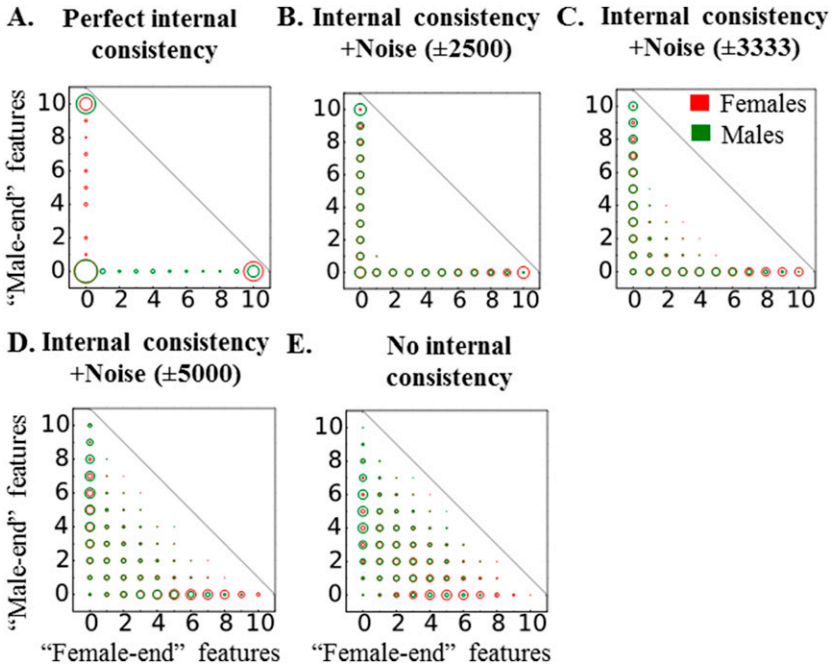

Fig. S1. Hypothetical distributions of "male-end" and "female-end" scores under different degrees of internal consistency. A bivariate scattergram of the number of regions at the "female-end" (*x* axis) and at the "male-end" (*y* axis) in females (red) and males (green) under (*A*) perfect internal consistency; (*B–D*) different degrees of noise (*B*) ±2,500, (*C*) ±3,333, (*D*) ±5,000; and (*E*) no internal consistency (see *SI Methods* and *SI Results* for details).

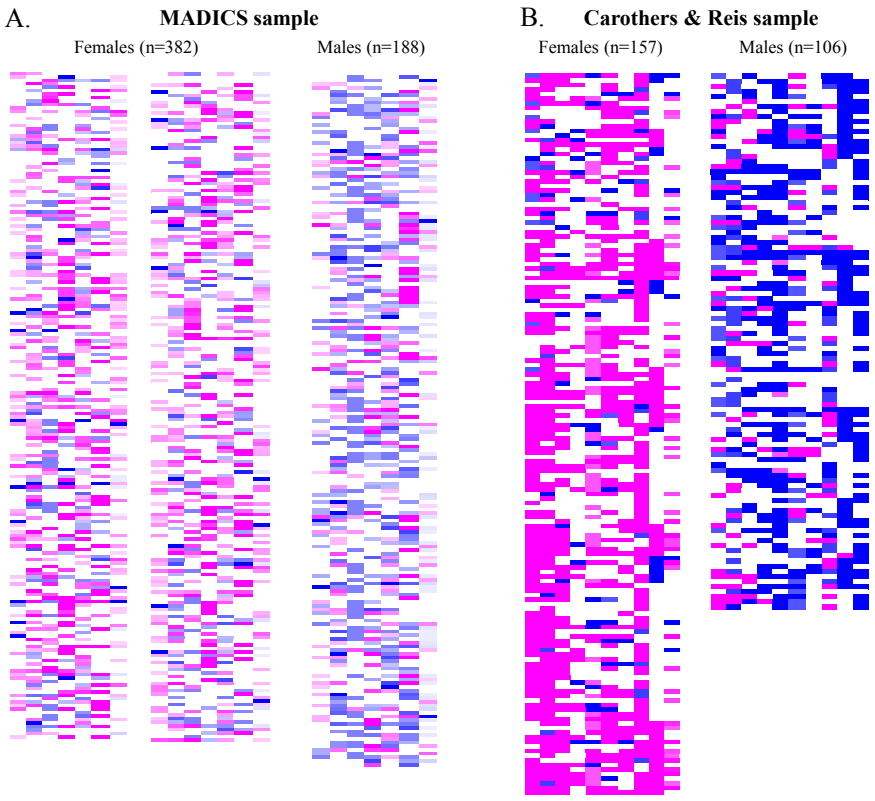

Fig. S2. The human gender mosaic. A color representation of the degree of femininity-masculinity [pink ("female-end") – white ("intermediate") – blue ("male-end")] of the variables showing the largest sex/gender differences in the (*A*) MADICS sample and (*B*) Carother and Reis' sample. The pink-white-blue scales were created as in Fig. 2*D*. Each line is a subject and each column is a single characteristic. The order of columns from left to right is the same as the order of variables in Table S5.

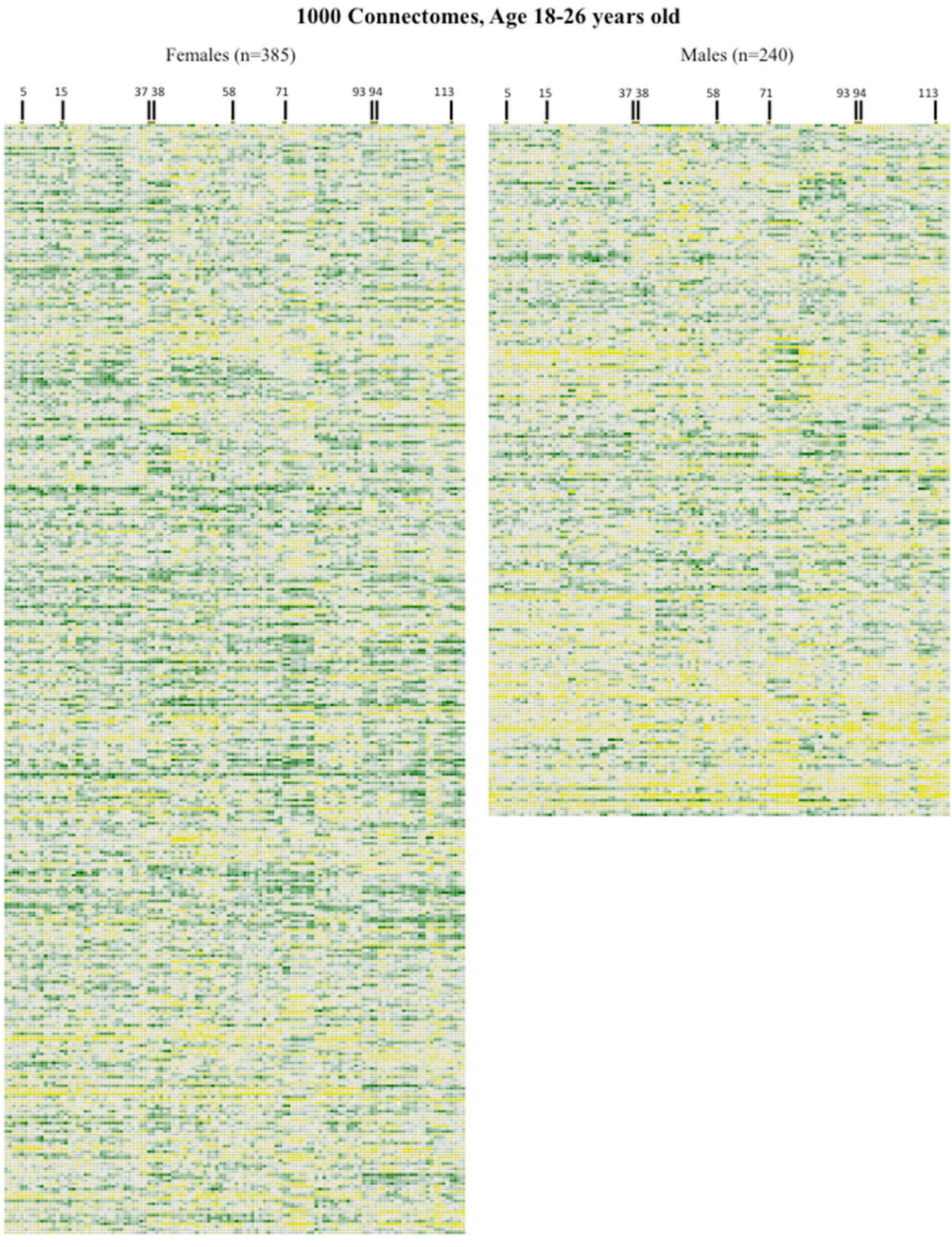

Fig. S3. The human brain mosaic. The gray matter volume of all 116 regions of gray matter in females (*Left*) and in males (*Right*) from the subsample of the 18–26 y olds from the 1000 Functional Connectomes Project sample is represented using a continuous high-low (green-white-yellow) scale, created as in Fig. 3. Each horizontal line represents the brain of a single subject and each column represents a single brain region. The regions that showed the largest sex/gender differences and were included in the internal consistency analysis are marked with a black bar. The number above each bar corresponds to the region's number in the AAL atlas and in Table S1.

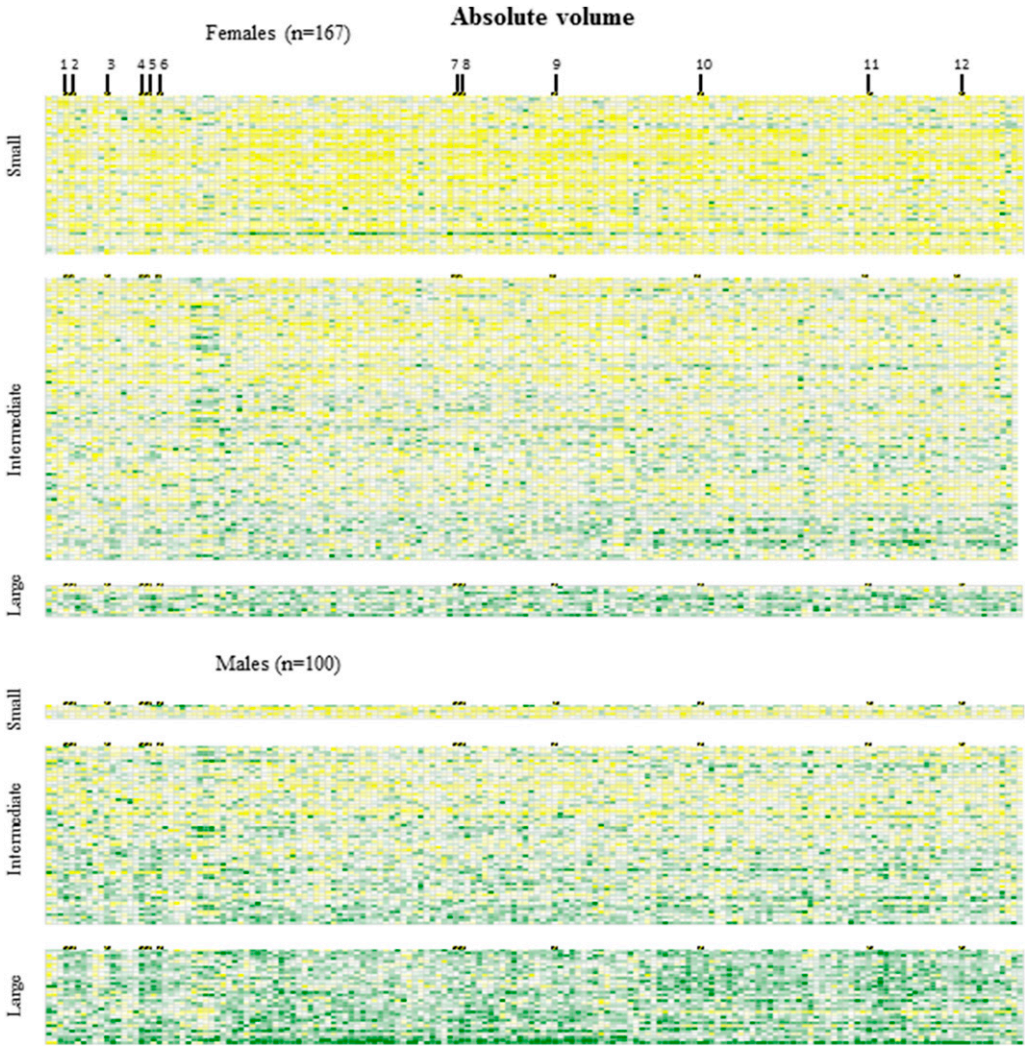

Fig. S4. The human brain mosaic. The absolute volume of 68 cortical regions, 23 subcortical regions, and 77 regions of white matter in females (*Upper*) and males (*Lower*) is represented using a continuous high-low (green-white-yellow) scale, created as in Fig. 3. The brains of both groups are separated into large ("male-end"), "intermediate", and small ("female-end"), according to total brain volume, to foster the appreciation of size effects and of sex/gender effects. Each horizontal line represents the brain of a single subject and each column represents a single brain region. The regions that showed the largest sex/gender differences and were included in the internal consistency analysis are marked with a black bar. The number above each bar corresponds to the region's number in Table S3.

Table S1. Size (Cohen's *d*) of the sex/gender difference in the brain regions included in the four analyses of internal consistency that used the AAL atlas

Region number	Standardized name	VBM first sample	VBM 1000* sample	VBM 1000* age 18–26 subsample	DTI FA
5	Left superior frontal gyrus, orbital part	0.81†	0.5	0.45†	−0.46
6	Right superior frontal gyrus, orbital part	0.84†	0.45	0.43	−0.44
13	Left inferior frontal gyrus, triangular part	0.52	0.44	0.37	−0.78†
15	Left inferior frontal gyrus, orbital part	0.45	0.49	0.46†	−0.51
25	Left superior frontal gyrus, medial orbital	0.77†	0.35	0.33	−0.15
27	Left gyrus rectus	0.83†	0.32	0.31	−0.46
28	Right gyrus rectus	0.75†	0.36	0.38	−0.32
37	Left hippocampus	0.74†	0.6†	0.46†	−0.06
38	Right hippocampus	0.72†	0.69†	0.58†	−0.23
57	Left postcentral gyrus	0.44	0.54†	0.42	−0.32
58	Right postcentral gyrus	0.51	0.56†	0.48†	−0.16
62	Right inferior parietal, excluding supramarginal and angular gyri	0.32	0.52†	0.43	−0.22
71	Left caudate nucleus	0.84†	0.6†	0.54†	−0.25
72	Right caudate nucleus	0.7†	0.47	0.37	−0.36
77	Left thalamus	−0.12	0.56†	0.37	−0.21
89	Left inferior temporal gyrus	0.06	0.08	0.07	−0.73†
93	Cerebellum left crus II	0.64	0.52†	0.46†	−0.58
94	Cerebellum right crus II	0.59	0.53†	0.51†	0
95	Cerebellum left hemispheric lobule III	0.42	0.2	0.06	−0.74†
97	Cerebellum left hemispheric lobule IV/V	0.52	0.37	0.25	−1.05†
98	Cerebellum right hemispheric lobule IV/V	0.61	0.29	0.16	−0.9†
99	Cerebellum left hemispheric lobule VI	0.33	0.51	0.39	−0.78†
100	Cerebellum right hemispheric lobule VI	0.35	0.42	0.3	−0.83†
101	Cerebellum left hemispheric lobule VIIB	0.66	0.46	0.43	−0.72†
103	Cerebellum left hemispheric lobule VIII	0.58	0.38	0.31	−0.74†
105	Cerebellum left hemispheric lobule IX	0.45	0.38	0.24	−0.84†
106	Cerebellum right hemispheric lobule IX	0.6	0.36	0.27	−0.81†
113	Vermic lobule VII	0.52	0.67†	0.6†	−0.49
116	Vermic lobule X	0.8†	0.24	0.14	−0.43

Positive values indicate that the average in females is larger than in males.
*1000, 1000 Functional Connectomes Project.
†This region was included in the analysis described in that column.

Table S2. Percent of brains showing substantial variability and percent of brains showing internal consistency under different definitions of the "male-end" and "female-end" zones

Dataset	Percent in end-zones			
	50%	33% (percent reported in the main text)	20%	10%
First sample, VBM	All: 66.5%, 6.4%	All: 34.5%, 2.4%	All: 10.0%, 0.7%	All: 2.1%, 0.0%
	♂: 65.2%, 11.6%	♂: 34.8%, 5.0%	♂: 10.7%, 1.8%	♂: 1.8%, 0.0%
	♀: 67.5%, 3.0%	♀: 34.3%, 0.6%	♀: 9.5%, 0.0%	♀: 2.4%, 0.0%
1000, VBM	All: 68.7%, 8.0%	All: 38.8%, 3.0%	All: 13.3%, 0.8%	All: 2.1%, 0.4%
	♂: 67.2%, 10.3%	♂: 36.9%, 5.3%	♂: 13.9%, 1.4%	♂: 1.7%, 0.6%
	♀: 69.7%, 6.3%	♀: 40.2%, 1.4%	♀: 12.9%, 0.4%	♀: 2.4%, 0.2%
1000, VBM 18–26 subsample	All: 78.9%, 4.6%	All: 53.4%, 1.1%	All: 23.2%, 0.2%	All: 5.0%, 0.0%
	♂: 77.1%, 3.8%	♂: 50.0%, 1.7%	♂: 22.1%, 0.3%	♂: 5.0%, 0.0%
	♀: 80.0%, 5.2%	♀: 55.6%, 0.8%	♀: 23.9%, 0.0%	♀: 4.9%, 0.0%
NKI, SBA, cortical thickness	All: 54.7%, 16.1%	All: 24.3%, 8.2%	All: 6.0%, 3.3%	All: 0.7%, 0.0%
	♂: 51.0%, 20.0%	♂: 21.0%, 10.0%	♂: 0.0%, 5.0%	♂: 0.0%, 0.0%
	♀: 56.9%, 13.8%	♀: 26.3%, 7.2%	♀: 9.6%, 2.4%	♀: 1.2%, 0.0%
NKI, SBA, volume of cortical and subcortical gray and white matter	All: 48.2%, 5.9%	All: 22.8%, 2.2%	All: 5.9%, 0.7%	All: 1.8%, 0.0%
	♂: 43.7%, 3.9%	♂: 25.2%, 1.9%	♂: 6.8%, 1.9%	♂: 2.9%, 0.0%
	♀: 50.9%, 7,1%	♀: 21.3%, 2.4%	♀: 5.3%, 0.0%	♀: 1.2%, 0.0%
DTI fractional anisotropy	All: 55.8%, 10.9%	All: 24.6%, 3.6%	All: 2.9%, 2.2%	All: 0.0%, 0.0%
	♂: 60.9%, 8.7%	♂: 29.0%, 2.9%	♂: 1.4%, 1.4%	♂: 0.0%, 0.0%
	♀: 50.7%, 13.0%	♀: 20.3%, 4.3%	♀: 4.3%, 2.9%	♀: 0.0%, 0.0%
DTI connectivity	All: 75.4%, 2.2%	All: 47.8%, 0.0%	All: 20.3%, 0.0%	All: 10.9%, 0.0%
	♂: 76.8%, 2.9%	♂: 52.2%, 0.0%	♂: 21.7%, 0.0%	♂: 15.9%, 0.0%
	♀: 73.9%, 1.4%	♀: 43.5%, 0.0%	♀: 18.8%, 0.0%	♀: 5.8%, 0.0%

Substantial variability: both "male-end" and "female-end" characteristics. Internal consistency: only "female-end" or only "male-end" characteristics (brains with only "intermediate" characteristics were not included here, because the narrower the end-zones, the wider the intermediate zone); percent in end-zones: percent of males and females with extreme scores included in the "male-end" and "female-end" zones, respectively; ♂, males; ♀, females.

Table S3. Size (Cohen's *d*) of the sex/gender difference in the brain regions included in the surface-based analysis

Number of region in Fig. S4	Standardized name	Cohen's *d*		
		Thickness	Volume gray matter	Volume white matter
	Left inferior temporal gyrus	−0.58*	−0.89	−0.94
10	Left medial orbital gyrus	−0.25	-0.94*	−0.82
	Left middle temporal gyrus	−0.56*	−0.93	−0.63
	Opercular part of left inferior frontal gyrus	−0.57*	−0.61	−0.51
	Left posterior cingulate gyrus	−0.45*	−0.71	−0.71
	Left insula	−0.47*	−0.78	−0.64
7	Right inferior parietal lobule	−0.06	−0.81	−0.97*
11, 8	Right inferior temporal gyrus	−0.48*	−0.95*	−0.94*
	Opercular part of right inferior frontal gyrus	−0.48*	−0.65	−0.54
12, 9	Right precuneus	−0.12	−0.95*	−0.99*
1	Left cerebellum	—	−1.03*	−0.8
2	Left thalamus - proper	—	−0.97*	—
3	Brainstem	—	−1.03*	—
4	Right cerebellum	—	−1.04*	−0.8
5	Right thalamus - proper	—	−0.99*	—
6	Right putamen	—	−0.94*	—

Positive values of Cohen's *d* indicate that the average in females is larger than in males.
*This region was included in the analysis described in that column.

Table S4. Size (Cohen's *d*) of the sex/gender difference in the strength of the connections included in the connectivity analysis

Connection between:	Cohen's *d*
Left superior temporal gyrus – left middle temporal gyrus	−0.96
Left calcarine fissure and surrounding cortex – left lingual gyrus	−0.83
Right anterior cingulate and paracingulate gyri – right middle cingulate and paracingulate gyri	−0.70
Left precentral gyrus – left angular gyrus	−0.69
Left supplementary motor area – left paracentral lobule	−0.67
Left superior frontal gyrus, orbital part – left inferior frontal gyrus, orbital part	−0.66
Right inferior frontal gyrus, orbital part – right insula	−0.66

Positive values of Cohen's *d* indicate that the average in females is larger than in males.

Table S5. Size (Cohen's *d*) of the sex/gender differences in the gender characteristics included in the three analyses

No. of column in figure	Name of variable	Number of items comprising the variable	Cronbach's α	Cohen's *d*
MADICS				
1	Expectations for sexual discrimination	3	0.68	0.57
2	Communication with mother	5	0.85	0.47
3	Worries about weight	4	0.91	0.78
4	Masculine self-esteem	3	0.68	−0.72
5	Communication with peers	6	0.84	0.44
6	Problem behavior	14	0.69	−0.43
7	Gender-related attitudes	17	0.56	0.65
ADD Health				
1	Perceived weight	1	—	0.44
2	Depression	11	0.80	0.41
3	Modified BEM femininity	12	0.90	0.52
4	Delinquency	12	0.70	−0.43
5	Impulsivity	9	0.87	−0.57
6	Gambling	1	—	−0.47
7	House work	2	0.69*	0.57
8	Engagement in sports	6	0.68	−0.47
Carothers and Reis				
1	Boxing	1	—	−1.17
2	Construction	1	—	−1.18
3	Playing golf	1	—	−1.01
4	Playing video games	1	—	−1.12
5	Scrapbooking	1	—	1.70
6	Taking a bath	1	—	1.01
7	Talking on the phone	1	—	1.11
8	Watching porn	1	—	−1.09
9	Watching talk shows	1	—	1.41
10	Cosmetics	1	—	2.17

Positive values of Cohen's *d* indicate that the average in females is larger than in males.
*Pearson correlation.