

**HUMAN BRAIN MAPPING****Machine Learning of Brain Gray Matter Differentiates Sex in  
a Large Forensic Sample**

Journal:	<i>Human Brain Mapping</i>
Manuscript ID	HBM-18-0372.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Anderson, Nathaniel; The Mind Research Network, Harenski, Keith; Mind Research Network Harenski, Carla; Mind Research Network, Koenigs, Michael; University of Wisconsin-Madison, Psychiatry Decety, Jean; The University of Chicago, Psychology Calhoun, Vince; Mind Research Network; University of New Mexico Kiehl, Kent; Mind Research Network; University of New Mexico
Keywords:	MRI, Source Based Morphometry, Antisocial Behavior, Classification, Gender, Support Vector Machine

SCHOLARONE™  
Manuscripts

**Machine Learning of Brain Gray Matter Differentiates Sex in a Large Forensic Sample**

Nathaniel E. Anderson <sup>a\*</sup>, Keith A. Harenski <sup>a</sup>, Carla L. Harenski <sup>a</sup>, Michael R. Koenigs <sup>b</sup>, Jean Decety <sup>c</sup>,  
Vince D. Calhoun <sup>a,d</sup>, Kent A. Kiehl <sup>a,d</sup>

- a. The Mind Research Network & Lovelace Biomedical and Environmental Research Institute,  
Albuquerque, NM, USA
- b. University of Wisconsin-Madison, Madison, WI, USA
- c. University of Chicago, Chicago, IL, USA
- d. University of New Mexico, Albuquerque, NM, USA
- Acknowledgements:** This research was supported in part by grants from the National Institute of Mental Health: R01MH070539, R01MH071896, R01MH085010, R01DA026505, R01DA026964, R01DA020870, R01EB005846, P20GM103472, and by The MacArthur Foundation Law and Neuroscience Project. We also thank the staff and inmates at New Mexico and Wisconsin Departments of Corrections; without their support, this research would not be possible.
- Conflict of Interest:** The authors declare that they have no conflict of interest.

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

2

**Abstract**

Differences between males and females have been extensively documented in biological, psychological, and behavioral domains. Among these, sex differences in the rate and typology of antisocial behavior remains one of the most conspicuous and enduring patterns among humans. However, the nature and extent of sexual dimorphism in the brain among antisocial populations remains mostly unexplored. Here we seek to understand sex differences in brain structure between incarcerated males and females in a large sample ( $n=1300$ ) using machine learning. We apply source-based morphometry (SBM), a contemporary multivariate approach for quantifying gray matter measured with magnetic resonance imaging (MRI), and carry these parcellations forward using machine learning to classify sex. Models using components of brain gray matter volume and concentration were able to differentiate between males and females with greater than 93% generalizable accuracy. Highly differentiated components include orbitofrontal and frontopolar regions, proportionally larger in females, and anterior medial temporal regions proportionally larger in males. We also provide a complimentary analysis of a non-forensic healthy control sample and replicate our 93% sex discrimination. These findings demonstrate that the brains of males and females are highly distinguishable. Understanding sex differences in the brain has implications for elucidating variability in the incidence and progression of disease, psychopathology, and differences in psychological traits and behavior. The reliability of these differences confirms the importance of sex as a moderator of individual differences in brain structure and suggests future research should consider sex specific models.

**Keywords:** Antisocial Behavior, Sex, Gender, MRI, Source Based Morphometry, Machine Learning

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Differences between males and females have been extensively documented in biological, psychological, and behavioral domains including criminology. These differences have encouraged the examination of sexual dimorphism and functional specialization in the human brain, which has remained a prominent branch of psychology and neuroscience research for well over a century (Broca, 1861; Lee & Pearson, 1901). Enduring controversy surrounds this topic as, historically, some have used the identification of organic differences among groups as a platform for reinforcing stereotypes and inferring a basis for fundamental limitations in ability, intelligence, and social status (Bean, 1906; Lynn, 1994; Russett, 2009). These findings and their interpretations have been rightfully challenged and debated for decades (Alper, 1985; Gould, 1996; Lynn, 1994). While developments in neuroimaging technology and computational methods have added new perspectives to the debate, they are no less prone to the controversies of *neurosexism* (Fine, 2013; Fine, Jordan-Young, Kaiser, & Rippon, 2013). Still, the elucidation of sex differences in the brain remains a highly important topic as these data help us understand more about the variable incidence and expression of mental health issues, disease, and our socio-biological and behavioral diversity as a species.

Despite the historically divisive nature of the topic, prevailing evidence at least does not favor a null hypothesis. The existence of organic differences in the brain, attributable to sex, has been well-replicated and reviewed at an aggregate level (Ruigrok et al., 2014; Sacher, Neumann, Okon-Singer, Gotowiec, & Villringer, 2013). Evidence suggests that men have larger brains than women, both absolutely and after controlling for average differences in body and head size (Allen, Damasio, Grabowski, Bruss, & Zhang, 2003; Lüders, Steinmetz, & Jäncke, 2002; Nopoulos, Flaum, O’Leary, & Andreasen, 2000). Morphologically, men’s brains have higher relative proportions of white matter while women’s brains have higher proportions of gray matter (Allen et al., 2003; Gur et al., 1999). Localized dimorphic regions are more variable in the literature, but accumulating evidence suggests that males

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

exhibit proportionally larger volumes in subcortical and interior cortical regions including the amygdala, hippocampus, parahippocampal gyrus, posterior cingulate, and temporal poles. Females, by contrast, exhibit proportionally larger volumes in several frontal cortical regions including inferior frontal gyrus, middle frontal gyrus, frontal pole, and frontal operculum (Ruigrok et al., 2014; Sacher et al., 2013).

Advances in the quantification of brain data have the potential to improve the sensitivity of these measures—for instance, the parcellation and boundaries of individual anatomical features often vary across studies, and methods that apply data driven approaches for defining these regions may provide additional benefits (Gupta et al., 2015; Xu, Groth, Pearlson, Schretlen, & Calhoun, 2009).

Recent descriptions have emphasized that despite the presence of well-replicated average differences in many brain regions between sexes, there is substantial overlap in the morphology of individual brain regions. Based on a sample of over 1400 individuals, Joel and colleagues (2014) argued that a lack of internal consistency in dimorphism on many individual metrics makes male and female brains essentially indistinguishable at an individual level. Critics, however, responded that these conclusions rely on univariate logic (Del Giudice et al., 2015, 2016; Rosenblatt, 2016), and others have demonstrated that multivariate classification techniques relying on variation among many brain regions simultaneously are quite reliable. For example, relying on a sample of over 1500 individuals, Chekroud, Ward, Rosenberg, & Holmes (2016) achieved multivariate classification of sex based on gray matter with over 90% accuracy (see also Lao et al., 2004).

**Sex Differences and Antisocial Behavior**

Sex differences in the brain also covary with many observable differences in behavior. A largely unexplored segment of this research lies in extending this work to better understand antisocial behavior in forensic samples. Conspicuous sex differences are apparent in both the frequency and nature of antisocial behavior (Heidensohn & Silvestri, 2012; Moffitt, 2001). Men engage in antisocial behavior more frequently than females (Del Giudice, 2015; Rowe, Vazsonyi, & Flannery, 1995), and these

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

differences are even more pronounced when focusing on aggressive behavior and violent crime (Archer, 2004). For example, approximately 75% of those arrested in the United States are males, and 80% of those arrested for violent crimes are males (Federal Bureau of Investigation, 2014). A number of theories examine different influences on these persistent trends from the perspectives of sociology (Harris, 2000; Steffensmeier & Allan, 1996), psychology (Bettencourt & Miller, 1996; Del Giudice, 2015; Feder, Levant, & Dean, 2007), and biology (Batrinos, 2012; Ferguson & Beaver, 2009). Neuroscience may have a unique place among these perspectives if we consider that, regardless of the variety of antecedents, and their interactions, these influences likely converge in the brain as it is the organ that governs our behavior most proximally (see also Jorgensen, Anderson, & Barnes, 2016). For example, Raine, Yang, Narr, and Toga (2011) reported that reductions in orbitofrontal and middle frontal gray matter volume accounted for approximately 77% of the difference in antisocial behavior between males and females. Furthermore, relationships between brain variables and antisocial behavior differentiated by sex emerge quite early in development, with boys and girls showing unique associations between brain measures and delinquent conduct (Michalska, Decety, Zeffiro, & Lahey, 2015; Michalska, Zeffiro, & Decety, 2016; Raschle et al., 2018;). In short, sex differences in the brain appear to be important for understanding differences in antisocial behavior; however, it has yet to be demonstrated whether such differences amount to simple average differences or more precisely represent complex dimorphism between sexes in an antisocial sample.

Here we set out to examine the utility of morphological gray matter differences for differentiating between male and female incarcerated offenders. We further aim to demonstrate the utility of an advanced multivariate approach for quantifying gray matter (source-based morphometry) and to apply machine learning algorithms to these data to examine their separability based on sex. These data will be influential for understanding the nature of sex differences in the brain and a step toward further clarifying their role in large-scale behavioral patterns in our society.

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

6

**Methods****Participants**

Data were collected from adult volunteers incarcerated in prisons in New Mexico and Wisconsin, and included male and female adult and juvenile offenders. A total of 1300 participants (males  $n = 1014$ ; females  $n = 286$ ) were included in this study. This sample was restricted to right-handed individuals. Ages ranged from 12 – 66 ( $M = 31.3$ ,  $SD = 10.8$ ). IQ ranged from 71 – 140 ( $M = 95.8$ ,  $SD = 13.1$ ); volunteers with  $IQ < 70$  were excluded from participation. Males and females did not differ on age or IQ, see sample characteristics in Table 1. Age and IQ were both used as covariates in prediction models. To observe best practices in prediction modeling approaches, this sample was randomly divided into two separate samples: one designated for training the machine ( $n = 930$ ), and the other sample dedicated for testing generalizability of the model ( $n = 370$ ), see classification methods below.

These data have been aggregated from participants who have volunteered for one of several ongoing brain imaging studies taking place in our collaborating forensic institutions. All research protocols have been approved by institutional review boards at the Ethical and Independent Review Services (E&I IRB) and University of Wisconsin-Madison and the Office of Human Research Protections (OHRP). This research meets ethical standards for responsible research with human subjects. Volunteers from forensic facilities were recruited via fliers and word of mouth. Meetings were scheduled with interested participants, and informed consent was obtained from all adult participants included in the study. Juvenile participants provided written informed assent in addition to their parent/guardian's written informed consent. Participants were informed of their right to terminate participation at any point, the lack of any direct institutional benefits, and that their participation would not affect their facility status or parole status. They were compensated with an hourly rate commensurate with standard rate for work assignments at their facility. Participants were screened for MRI safety, and

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

excluded for any contraindications (e.g. metal in body). Participants were also excluded from our studies if they reported a history of major head injury.

**MRI Data Acquisition and Analysis**

High-resolution T1-weighted structural MRI scans were acquired using the Mind Research Network Mobile Siemens 1.5T Avanto MRI scanner. This custom unit is specifically designed for transport and deployment within the secure environments of forensic facilities. A multi-echo MPAGE pulse sequence (repetition time=2530ms, echo times=1.64ms, 3.50ms, 5.36ms, 7.22ms, inversion time=1100ms, flip angle=7°, slice thickness=1.3mm, matrix size=256×256) was used, yielding 128 sagittal slices with an in-plane resolution of 1.0mm×1.0mm. Data were pre-processed and analyzed using Statistical Parametric Mapping software (SPM12; <http://www.fil.ion.ucl.ac.uk/spm>). T1 images were manually inspected by an operator blind to subject identity and realigned to ensure proper spatial normalization. Images were then analyzed via the unified segmentation approach as implemented in SPM12 (Ashburner & Friston, 2005). Unified segmentation allows for image registration based on Gaussian mixture modelling, tissue classification with warped prior probability maps and bias correction to be combined in the same generative model. Both volume (modulated data) and density (unmodulated data) were extracted for analyses. A Jacobian modulation was performed to preserve total volume (Ashburner & Friston, 2000; 2005). A nonlinear transformation without Jacobian determinants was performed on unmodulated images to extract grey matter density (Ashburner & Friston, 2000; 2005). Modulated and unmodulated images were resampled to 2x2x2 mm and smoothed with a 10mm full-width at half-maximum (FWHM) Gaussian kernel. Voxels with grey matter value of <0.15 were excluded in order to remove possible edge effects between grey matter and white matter. Only grey matter segments were used in this analysis.

Source Based Morphometry (SBM) was used to separate gray matter volume and density into maximally independent *source networks*. SBM is a multivariate alternative to voxel-based morphometry



## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

(VBM). It is an application of Independent Component Analysis (ICA) for structural brain data (Calhoun & Adali, 2006; Xu et al., 2009). In contrast to VBM, SBM uses relationships among voxels, identifying distinct regions with common covariation between subjects. These source networks are spatially distinct and maximally independent (see also Caprihan et al., 2011; Nakai et al., 2004). SBM was carried out using spatial ICA as implemented in the GIFT Toolbox (Calhoun, Liu, & Adali, 2009) (<http://mialab.mrn.org/software/gift>). The number of source networks for modulated and unmodulated gray matter was set at 30 each, which has been demonstrated to produce a stable set of structural networks (Turner et al., 2012; Xu et al., 2009). Final component selection included the removal of spatial maps that were identified as noise or motion related. One modulated and two unmodulated components were removed, yielding 57 independent components used in the final analysis. Loading coefficients for each source network for each individual are carried forward as features in the classification models described below.

**Machine Learning Classification**

Loading coefficients for modulated and unmodulated gray matter from SBM analysis were used together with age and IQ estimates to test their utility in classifying sex (as categorically male or female). A large number of data classification approaches are available, each with certain advantages. While the mathematical approaches differ, these machine learning methods serve essentially the same purpose: to use a set of known information to build a generalizable model capable of predicting unknown information from a set of cases with limited data. In order to test the utility of a given type of information (for our purposes, localized gray matter volumes and density) for differentiating among groups/classes (sex), we provide full information on a set of cases for deriving the initial model, then test the performance (generalizability) of that model on cases for which limited data are provided, but for which the true value of a predicted outcome is known, such that errors in classification can be detected (i.e. supervised learning). In order to examine classification performance under best practices

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

for reproducibility (Friedman, Hastie, & Tibshirani, 2001; Vapnik, 1998), we first separated our data into a training sample ( $n = 930$ ) and a testing sample ( $n = 370$ ), based on random computer-aided assignment. Individuals in the testing sample were not included in model development to avoid potential bias. Their data were solely used for the purpose of validating the generalizability of the classification performance of models developed on the training set.

To compare among a number of methods, we used a publicly available toolbox which runs multiple classification approaches simultaneously to explore their relative performance. The *Polyssifier* toolbox (<https://github.com/alvarouc/polyssifier>) compares a k-nearest neighbors, linear support vector machine, radial basis function (RBF) support vector machine (SVM), decision tree, random forest, logistic regression, naïve Bayes, and voting classifier approaches. For the present data, SVM and logistic regression approaches exhibited the highest accuracy in performance, as we have found in several other datasets examining classification using brain data from incarcerated samples (e.g. Cope et al., 2014; Kiehl et al., 2018; Steele et al. 2015; Steele, Rao, Calhoun, & Kiehl, 2017). Here we reserve our descriptions of methods and outcomes to RBF SVM and logistic regression. All classifications and model testing were run in the *R* software package using the CARET (for SVM) (Kuhn et al., 2016) and glmnet (for logistic regression) libraries (Friedman, Hastie, & Tibshirani, 2010).

In SVM classification (Friedman et al., 2001; Vapnik, 1998) the goal is to construct a hyperplane separating the training examples within the feature space into two subspaces. The classification procedure takes place in two stages: training and testing. During training the algorithm finds a hyperplane (i.e. a high dimensional plane) that maximizes the margin between the training samples of both classes. The training samples that lie on the margins of the hyperplane are the called support vectors. These are the units/participants most difficult to classify and therefore define the boundaries of the hyperplane. Kernel functions can be applied if the data is not linearly separable in the original space and allows for group classification based on non-linear effects, e.g. RBF SVM. Once the decision

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

10

boundary is learned from the training data, it can be used to predict the class of new test samples.

Subject classification and best parameter selection via grid search over a reasonable range of values was accomplished using five repetitions of 10-fold cross validation in *R*.

For comparison, we also implemented an *elastic net* penalized logistic regression model. This method is designed to classify data based on sets of highly correlated predictors. It combines the properties of LASSO and ridge regression for promoting automatic variable selection while reducing the impact of highly correlated predictors. This method is theoretically appropriate due to the correlated nature of brain data within individuals (Bunea et al., 2011). Both modulated and unmodulated data from gray matter were used for prediction as both measures may be incrementally useful for differentiation of individual differences (e.g. Meda et al. 2008). For both SVM and logistic regression classification methods, models were developed on the training sample ( $n=930$ ) and then applied, unchanged, to the hold-out testing sample ( $n=370$ ).

From 30 modulated and unmodulated components derived from SBM, 3 components (1 modulated, 2 unmodulated) were determined to be noise and/or motion-related, and these were removed from further analyses. Each classification method then utilized brain data from 57 (29 modulated, 28 unmodulated) components. Implementation of a feature selection step did not improve overall performance of these models. Likewise, simple exclusion of any individual components, list-wise, did not significantly reduce prediction accuracy, suggesting all components are simultaneously, and incrementally useful in the resulting models. In other words, no single component can be meaningfully identified as the most important to classification, apart from other features of the model. In order to select and illustrate component maps (brain areas) contributing to classifications, we performed independent samples t-tests comparing males and females using the loading coefficients on each of the 57 components from SBM. Modulated and unmodulated components showing the highest *t*-values for *males > females* and *females > males* are presented for illustration purposes; however, their

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

relative value in prediction models requires the simultaneous inclusion of all components utilized in the final models. Individuals' SVM loadings are dimensional and scale proportionally with gray matter volume/density, but loadings are not directly observable as discrete differences between groups.

To illustrate these differences, we have also carried out two-sample t-tests on gray matter volume (modulated VBM data) for the full sample of incarcerated males and females (n = 1300), and a non-incarcerated comparison group. This provides a more direct visualization of discrete gray matter differences between males and females.

**Non-Incarcerated Comparison Sample**

In order to compare and demonstrate the utility of the described SBM method for classifying sex in a non-incarcerated sample, we applied the same methods described above to the publicly available dataset from the Brain Genomics Superstruct Project ([www.neuroinfo.org/gsp/](http://www.neuroinfo.org/gsp/)). We restricted this sample to right-handed individuals only, yielding a total of n = 1448 (male/female). This sample was split randomly into training (n = 922; 550 female) and testing (n = 526; 296 female) sets. A total of 30 modulated and 30 unmodulated gray matter components were defined via SBM. Two modulated and two unmodulated components were determined to be noise yielding a total of 56 total components (28 modulated, 28 unmodulated) carried into the final prediction models. Mirroring analyses above, SVM and penalized logistic regression were carried out to classify males and females.

**Results**

Classification tables (training and testing samples) are provided for both SVM and elastic net logistic regression (see Tables 2-5). Cross-validated training samples yielded overall classification accuracy at 98.6% and 96.7% for SVM and logistic regression, respectively. Testing validation on the naïve hold-out set revealed generalizable overall classification accuracy at 93.8% for SVM and 93.0% for logistic regression.

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

12

In order to illustrate the structural components that are used as features in these classifications, independent samples *t*-tests comparing males and females on individuals' loading coefficients in SPSS v20 were carried out on all 57 (29 modulated, 28 unmodulated) SBM components. A selection of components demonstrating significant differences between males and females is provided in Figures 1 & 2. These component maps show the spatial organization of a selection of four (out of 57 total) structural networks that were useful in the classification models. These four were selected by carrying out *t*-tests between males and females in the training sample ( $n = 930$ ) on the loading coefficients for all components. For illustrative purposes components in loadings for gray matter volume (modulated; Figure 1) and concentration (unmodulated; Figure 2) were larger for males (blue) and females (orange) are presented. These maps do not represent volumetric differences, per se, between men and women, but simply the organization of structural components across the entire sample. Increasing volume or density in the highlighted regions is distinguishing of males and females in the classification of this sample. A component that is significantly larger for females, for example indicates that females have larger loading coefficients for that component than males, and these differences (across 57 components) were ultimately useful for classification. These loading coefficients thus scale proportionally with volumetric differences, which can be directly visualized with two-sample *t*-tests of VBM data.

**VBM Comparison**

*T*-tests directly comparing gray matter volume and density between males and females are shown in Figure 3. These brain maps summarize voxel-by-voxel statistics across the entire brain illustrating areas with relatively higher gray matter volume in males and females. SBM component differences and *t*-test comparisons confirm prominent regional differences include orbitofrontal and frontopolar regions, proportionally larger in females, and anterior/medial temporal regions, proportionally larger in males (see Figure 3).

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

**Non-Incarcerated Comparison Sample**

SBM-based classification results for the non-incarcerated comparison sample were virtually identical to those from the incarcerated sample. Classification on the naïve testing data (n = 526) produced overall accuracy of approximately 93.5% for SVM and 94.1% for penalized logistic regression methods. Confusion matrices for these methods are provided in Tables 6 & 7.

**Discussion**

The purpose of this study was twofold. First, we set out to examine the reliability of sexual dimorphism in the brain among a large incarcerated sample, extending work carried out by others in non-incarcerated healthy control subjects (c.f. Chekroud et al., 2016). Second, we aimed to demonstrate the utility of an ICA-based quantification method of gray matter for predicting sex from brain data, rather than relying on pre-established anatomical divisions of the brain. As expected, our results demonstrate that sexual dimorphism in brain structure is highly apparent among incarcerated samples, and the multivariate methods used to quantify gray matter allowed greater than 93% accuracy in classifying individuals as male or female. These findings are an important incremental step for elucidating neurobiological influences on male and female criminal behavior, underscoring the importance of sex differences in antisocial brains. Further, this is the first demonstration of brain-based sex classifications in an incarcerated sample, demonstrating that multivariate patterns of gray matter are reliably dimorphic between sexes among male and female inmates. Finally, to our knowledge this is the first report to demonstrate that ICA-based parcellation of brain anatomy is as effective, if not more reliable, than traditional atlas-based approaches or voxel-based morphometry for defining sex-specific differences.

The implications of this work are multifaceted and stand to encourage more precise scientific work aimed at understanding of the biological influences on crime. Our data suggest that the brain

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

14

structure of antisocial males and females are no less distinguishable than what has been demonstrated among normative samples. This supports the notion that antisocial behavior in women should also be examined independently from conclusions that have been drawn from a traditionally male-focused literature. It also emphasizes the importance of considering biological-behavioral patterns separately in each gender, or at least considering the moderating effects of sex in these outcomes.

More generally, our findings challenge recent suggestions that male and female brains are essentially indistinguishable (Joel et al., 2015). The present findings add to a number of demonstrations that multivariate patterns in brain anatomy are quite reliable at classifying individual male and female brains (Chekroud et al., 2016; Del Giudice et al., 2015, 2016; Lao et al., 2004). While we agree that considerable overlap exists in the distribution of individual brain measures, concluding their effective equivalence misrepresents the complexity of the issue. The consequence of this would be to mistakenly undermine the value of drawing on sex differences to better understand individual differences in healthy and disordered brain function and behavior, including important differences in the incidence and progression of disease and psychopathology.

We suspect that conclusions endorsing the uniformity of male and female brains are at least partially undergirded by a benevolent motivation to subvert sexist attitudes. As noted above, neuroanatomical sexual dimorphism is a perpetually controversial topic since it has often been used as a defense of prejudiced beliefs, sex-based stereotypes, and categorical limitations in abilities. Further, it has often been misappropriated to justify unfair or unequal treatment in social domains. What should be clear, however, is that sexist interpretations as well as the notion of equivalence between sexes both misconstrue and undermine the true value of understanding sex differences. The implicit presumption in both of these positions seems to be that *different* must mean *unequal*, branding one as *inferior*, but males and females do not need to be identical to be treated equally. Further, developing a better understanding of the origins and consequences of those differences may be imperative to continued

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

progress in behavioral neuroscience. Understanding the extent and boundaries of sex-based differences in the brain will ultimately help us better interpret differences in mental health issues, behavior, mortality, and the fundamental causes and contextual limitations of these differences.

The other important aim of this study was to demonstrate that a multivariate, data-driven technique for parceling independent components of brain anatomy (ICA) is an effective method for quantifying these sex-based differences. While sexual dimorphism in the brain has been studied for over a century, we are no longer limited to weighing post-mortem brains, as Broca did in the late 19<sup>th</sup> century (Broca, 1861), or inferring in vivo brain size from cranial circumference (Lee & Pearson, 1901; Rushton, 1992). Major historical innovations including magnetic resonance imaging have been coupled with increasingly sensitive analytic methods, like those applying ICA to MRI data. Improving the techniques and testing novel ways of quantifying brain data are instrumental steps in improving the utility of brain data for practical purposes. Comparing our data with studies that have used automatic parcellation, we demonstrate comparable, and even slightly improved classification accuracy. This may be the result of the ICA methods partitioning variance in a way that is optimized for quantifying individual differences in a way that increases sensitivity (Allen, Erhardt, Wei, Eichele, & Calhoun, 2012; Calhoun, Pearlson, & Pekar, 2001).

Components derived from ICA demonstrate voxels of gray matter that vary systematically with one another across subjects. The loading coefficients from the ICA solution are what successfully classify individuals as males and females in this demonstration. These loading coefficients indicate the relative contribution of an individual to the ICA solution that gave rise to that component. Thus, individuals with smaller absolute loading coefficients for a given component are less important for identifying and quantifying that network. The illustrative components provided in Figures 1 and 2 should not be interpreted directly as volumetric differences between males and females. Instead, values in each component are indicative of a pattern of gray matter structure that is more discriminative



## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

16

of males (blue) and females (orange) in our sample. Still, these loadings will scale proportionally with gray matter volume, and t-tests comparing volume directly across sex are provided for ease of viewing.

Anatomically, Figure 1 shows a component of increasing gray matter volume around the amygdala and anterior temporal cortex that is discriminating of males (blue area), while increased temporoparietal volume (BA40) extending into somatosensory areas is more discriminating of females (orange areas). Figure 2 shows that increased density in the anterior insula and medial prefrontal cortex is discriminating of males, while increasing density in the frontopolar and ventromedial/orbitofrontal gray matter is more discriminating of females. T-test maps (Figure 3) illustrate how these regions are systematically reflected in volumetric comparisons of males and females as well.

Importantly, brain regions differing between males and females have conspicuous roles in many cognitive features that align with previously identified patterns relevant to antisocial behavior. The amygdala plays a well-known role in detecting threat and in fear conditioning and coordination of responses to threat (Wilensky, Schafe, Kristensen, & LeDoux, 2006). Exaggerated activity in the amygdala and diminished orbitofrontal activity has been associated with patterns of reactive aggression and violence (Coccaro, McCloskey, Fitzgerald, & Phan, 2007; Meyer-Lindenberg et al., 2006). The anterior temporal cortex (temporal pole, BA38) is tightly connected to limbic and paralimbic structures influencing complex social and emotional processing (Olson, Plotzker, & Ezzyat, 2007). Lower volume and functioning in these regions have also been associated with diminished inhibition of violent and aggressive behavior (Blair, 2016; Coccaro et al., 2007; Tiihonen et al., 2008). The frontopolar and orbitofrontal regions are important in moral judgment and planning/executing behavior, including incorporating contingencies that involve consequences and other learned reinforcement (Greene, & Haidt, 2002; Moll, Oliveira-Souza, Bramati, & Grafman, 2002). The temporoparietal junction is also fundamental for execution of attentional shifts required for perspective-taking, theory of mind, and empathy (Decety & Lamm, 2007). It is noteworthy, however, that these prevailing patterns in antisocial

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

and externalizing populations have been studied almost exclusively among males. Still, it remains reasonable to suspect that higher relative rates of antisocial behavior, and particularly violent-antisocial behavior, in males may be tied to sex-specific differences in these brain areas (cf. Raine et al. 2011).

While on-average differences (cf. Figure 3) in these regions have been recognized previously in healthy samples (Ruigrok et al., 2014; Sacher et al., 2013), it is conceptually important that the current work has demonstrated multivariate models that approach something closer to truly dimorphic patterns (see McCarthy et al. 2012), effectively differentiating individuals into categorical groups based on intrinsic structural networks (via SBM components). Demonstrating these patterns highlights that antisocial behavior in men and women occurs in the context of unique neural architecture that accompanies highly reliable dimorphic sex differences. It further underscores the importance of accounting for sex differences when conducting cognitive neuroscience research among forensic samples.

**Limitations, future directions, and conclusions**

The results of the present study are accompanied by a number of limitations that should promote additional research on these important topics. Our study examines a relatively simple model intended to account primarily for the influence of biological sex on volumetric brain data, but many other moderating variables and quantitative methods are likely to improve our understanding of this topic. The present descriptions of sexual dimorphism in the human brain remain relatively macroscopic and imprecise. Future studies should examine discriminability based on measures that include functional activity as well as structural and functional connectivity (e.g. Ingallhalikar et al., 2014; Yaesoubi et al., 2015). Sex-related differences in brain structure and function are likely to interact with age as well (Brain Development Cooperative Group, 2012; Lenroot et al., 2007). In this study, we used age as a covariate in our models, but future studies may examine how this discriminability changes more precisely as a function of developmental progress over the lifespan (see also Kiehl et al., 2018).

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

18

The present study examined sexual dimorphism among incarcerated offenders primarily to establish essential discriminability and underscore the value of considering the moderating effects of sex/gender in forensic neuroscience. We ardently condemn the historically slippery perspective of relying on established brain differences to justify inequitable treatment based on prejudice. We simultaneously caution against the conclusion that male and female brains are indistinguishable, as this has the potential to undermine the real utility of these data in understanding our diversity as a species. It should be clear from these data that males and females exhibit morphometric differences that are reliable enough to predict sex with high individual specificity. The way we choose to utilize this information can either improve ongoing scientific rigor or alternatively reinforce obsolete ideas and stagnate progress.

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

References

Adler, F. S., & Adler, H. M. (1975). *Sisters in crime: The rise of the new female criminal*. New York: McGraw-Hill.

Allen, E. A., Erhardt, E. B., Wei, Y., Eichele, T., & Calhoun, V. D. (2012). Capturing inter-subject variability with group independent component analysis of fMRI data: a simulation study. *Neuroimage*, 59(4), 4141-4159. doi: 10.1016/j.neuroimage.2011.10.010

Allen, J. S., Damasio, H., Grabowski, T. J., Bruss, J., & Zhang, W. (2003). Sexual dimorphism and asymmetries in the gray–white composition of the human cerebrum. *Neuroimage*, 18(4), 880-894. doi: 10.1016/S1053-8119(03)00034-X

Alper, J. S. (1985). Sex differences in brain asymmetry: A critical analysis. *Feminist Studies*, 11(1), 7-37. doi: 10.2307/3180130

Archer, J. (2004). Sex differences in aggression in real-world settings: a meta-analytic review. *Review of General Psychology*, 8(4), 291-322. doi: 10.1037/1089-2680.8.4.291

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, 9(3), 212-230. doi: 10.1207/s15327957pspr0903\_2

Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry--The Methods. *Neuroimage*, 11(6), 805-821. doi: 10.1006/nimg.2000.0582

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage*, 26(3), 839-851. doi: 10.1016/j.neuroimage.2005.02.018

Batrinou, M. L. (2012). Testosterone and aggressive behavior in man. *International journal of endocrinology and metabolism*, 10(3), 563-568. doi: 10.5812/ijem.3661

Bean, R. B. (1906). Some racial peculiarities of the Negro brain. *Developmental Dynamics*, 5(4), 353-432. doi: 10.1002/aja.1000050402

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

20

- Bettencourt, B., & Miller, N. (1996). Gender differences in aggression as a function of provocation: a meta-analysis. *Psychological Bulletin*, 119(3), 422-447. doi: 10.1037/0033-2909.119.3.422
- Blair, R. J. (2016). The neurobiology of impulsive aggression. *Journal of child and adolescent psychopharmacology*, 26(1), 4-9. doi: 10.1089/cap.2015.0088
- 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. *Cerebral Cortex*, 22(1), 1-12. doi: 10.1093/cercor/bhr018
- Broca, P. (1861). *Sur le volume et la forme du cerveau suivant les individus et suivant les races* (Vol. 1). Harvard University: Hennuyer.
- Bunea, F., She, Y., Ombao, H., Gongvatana, A., Devlin, K., & Cohen, R. (2011). Penalized least squares regression methods and applications to neuroimaging. *Neuroimage*, 55(4), 1519-1527. doi: 10.1016/j.neuroimage.2010.12.028
- Calhoun, V. D., & Adali, T. (2006). Unmixing fMRI with independent component analysis. *IEEE Engineering in Medicine and Biology Magazine*, 25(2), 79-90. doi: 10.1109/MEMB.2006.1607672
- Calhoun, V. D., Liu, J., & Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage*, 45(1), S163-S172. doi: 10.1016/j.neuroimage.2008.10.057
- Calhoun, V. D., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3), 140-151. doi: 10.1002/hbm.1048
- Campbell, A. (1981). *Girl Delinquents*. UK: Oxford: Basil Blackwell.
- Caprihan, A., Abbott, C., Yamamoto, J., Pearlson, G., Perrone-Bizzozero, N., Sui, J., & Calhoun, V. D. (2011). Source-based morphometry analysis of group differences in fractional anisotropy in schizophrenia. *Brain Connectivity*, 1(2), 133-145. doi: 10.1089/brain.2011.0015

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

- Carrington, K. (2006). Does feminism spoil girls? Explanations for official rises in female delinquency. *Australian & New Zealand Journal of Criminology*, 39(1), 34-53. doi: 10.1375/acri.39.1.34
- Chekroud, A. M., Ward, E. J., Rosenberg, M. D., & Holmes, A. J. (2016). Patterns in the human brain mosaic discriminate males from females. *Proceedings of the National Academy of Sciences*, 113(14), E1968-E1968. doi: 10.1073/pnas.1523888113
- Chesney-Lind, M., Morash, M., & Stevens, T. (2008). Girls troubles, girls' delinquency, and gender responsive programming: A review. *Australian & New Zealand Journal of Criminology*, 41(1), 162-189. doi: 10.1375/acri.41.1.162
- Coccaro, E. F., McCloskey, M. S., Fitzgerald, D. A., & Phan, K. L. (2007). Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. *Biological psychiatry*, 62(2), 168-178. doi: 10.1016/j.biopsych.2006.08.024
- Cope, L. M., Ermer, E., Gaudet, L. M., Steele, V. R., Eckhardt, A. L., Arbabshirani, M. R., ... & Kiehl, K. A. (2014). Abnormal brain structure in youth who commit homicide. *Neuroimage: clinical*, 4, 800-807. doi: 10.1016/j.nicl.2014.05.002
- Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *The Neuroscientist*, 13(6), 580-593. doi: 10.1177/1073858407304654
- Del Giudice, M. (2015). Gender differences in personality and social behavior. In J. D. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences* (2nd ed., pp. 750-756). New York, NY: Elsevier.
- Del Giudice, M., Lipka, R. A., Puts, D. A., Bailey, D. H., Bailey, J. M., & Schmitt, D. P. (2015). Mosaic Brains? A Methodological Critique of Joel et al. (2015). Retrieved from: [http://cogprints.org/10046/1/Delgiudice\\_etal\\_critique\\_joel\\_2015.pdf](http://cogprints.org/10046/1/Delgiudice_etal_critique_joel_2015.pdf)

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

22

- Del Giudice, M., Lippa, R. A., Puts, D. A., Bailey, D. H., Bailey, J. M., & Schmitt, D. P. (2016). Joel et al.'s method systematically fails to detect large, consistent sex differences. *Proceedings of the National Academy of Sciences*, 113(14), E1965-E1965. doi: 10.1073/pnas.1525534113
- Federal Bureau of Investigation (2014). *Crime in the United States*. from <https://ucr.fbi.gov/crime-in-the-u.s/2014/crime-in-the-u.s.-2014/persons-arrested/main>
- Feder, J., Levant, R. F., & Dean, J. (2007). Boys and violence: A gender-informed analysis. *Professional Psychology: Research and Practice*, 38(4), 385-391. doi: 10.1037/0735-7028.38.4.385
- Ferguson, C. J., & Beaver, K. M. (2009). Natural born killers: The genetic origins of extreme violence. *Aggression and Violent Behavior*, 14(5), 286-294. doi: 10.1016/j.avb.2009.03.005
- Fine, C. (2013). Is there neurosexism in functional neuroimaging investigations of sex differences? *Neuroethics*, 6(2), 369-409. doi: 10.1007/s12152-012-9169-1
- Fine, C., Jordan-Young, R., Kaiser, A., & Rippon, G. (2013). Plasticity, plasticity, plasticity... and the rigid problem of sex. *Trends in Cognitive Sciences*, 17(11), 550-551. doi: 10.1016/j.tics.2013.08.010
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The Elements of Statistical Learning* (Vol. 1). New York: Springer.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). GLMNET: Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1-22.
- Gould, S. J. (1996). *The Mismeasure of Man*. New York, NY: WW Norton & Company.
- Greenfeld, L. A., & Snell, T. L. (2000). *Women Offenders*. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Gupta, C. N., Calhoun, V. D., Rachakonda, S., Chen, J., Patel, V., Liu, J., ... & Buitelaar, J. (2014). Patterns of gray matter abnormalities in schizophrenia based on an international mega-analysis. *Schizophrenia bulletin*, 41(5), 1133-1142. doi: 10.1093/schbul/sbu177

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

- Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. *Journal of Neuroscience*, 19(10), 4065-4072.
- Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work?. *Trends in cognitive sciences*, 6(12), 517-523. doi: 10.1016/S1364-6613(02)02011-9
- Harris, A. P. (2000). Gender, violence, race, and criminal justice. *Stanford Law Review*, 777-807.
- Heidensohn, F. (1989). Gender and Crime. In *Crime and Society: Sociology for a Changing World* (pp. 85-111). London: Palgrave, Springer. doi: 10.1007/978-1-349-19763-7\_5
- Heidensohn, F., & Silvestri, M. (2012). Gender and Crime. In M. Maguire, R. Morgan & R. Reiner (Eds.), *The Oxford Handbook of Criminology*. Oxford, UK: Oxford University Press.
- Heimer, K. (Producer). (2000). Changes in the gender gap in crime and women's economic marginalization. *The Nature of Crime: Continuity and Change*. Retrieved from [https://www.ncjrs.gov/criminal\\_justice2000/vol\\_1/02i.pdf](https://www.ncjrs.gov/criminal_justice2000/vol_1/02i.pdf)
- Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., . . . Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences*, 111(2), 823-828. doi: 10.1073/pnas.1316909110
- Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., . . . Margulies, D. S. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences*, 112(50), 15468-15473. doi: 10.1073/pnas.1509654112
- Jorgensen, C., Anderson, N. E., & Barnes, J. (2016). Bad Brains: Crime and Drug Abuse from a Neurocriminological Perspective. *American Journal of Criminal Justice*, 41(1), 47-69. doi: 10.1007/s12103-015-9328-0



## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

24

- Kiehl, K. A., Anderson, N. E., Aharoni, E., Maurer, J. M., Harenski, K. A., Rao, V., ... & Kosson, D. (2018). Age of gray matters: Neuroprediction of recidivism. *NeuroImage: Clinical*, 19, 813-823. doi: 10.1016/j.nicl.2018.05.036
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., . . . Hunt, T. (Producer). (2016). CARET: Classification and Regression Training, R package version 6.0-73. Retrieved from <https://cran.r-project.org/web/packages/caret/caret.pdf>
- Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S. M., & Davatzikos, C. (2004). Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*, 21(1), 46-57. doi: 10.1016/j.neuroimage.2003.09.027
- Lee, A. E., & Pearson, K. (1901). Data for the Problem of Evolution in Man: A First Study of the Correlation of the Human Skull. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical, and Engineering Sciences*, 196, 225-264. doi: 10.1098/rsta.1901.0005
- Lenroot, R. K., Gogtay, N., Greenstein, D. K., Wells, E. M., Wallace, G. L., Clasen, L. S., . . . Evans, A. C. (2007). Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage*, 36(4), 1065-1073. doi: 10.1016/j.neuroimage.2007.03.053
- Lüders, E., Steinmetz, H., & Jäncke, L. (2002). Brain size and grey matter volume in the healthy human brain. *NeuroReport*, 13(17), 2371-2374. doi: 10.1097/00001756-200212030-00040
- Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences*, 17(2), 257-271. doi: 10.1016/0191-8869(94)90030-2
- 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. *Journal of Neuroscience*, 32(7), 2241-2247. doi: 10.1523/JNEUROSCI.5372-11.2012
- Meda, S. A., Giuliani, N. R., Calhoun, V. D., Jagannathan, K., Schretlen, D. J., Pulver, A., ... & Sharma, T. (2008). A large scale (N= 400) investigation of gray matter differences in schizophrenia using

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

optimized voxel-based morphometry. *Schizophrenia research*, 101(1-3), 95-105. doi: 10.1016/j.schres.2008.02.007

Meyer-Lindenberg, A., Buckholtz, J. W., Kolachana, B., Hariri, A. R., Pezawas, L., Blasi, G., ... & Egan, M. (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proceedings of the National Academy of Sciences*, 103(16), 6269-6274. doi: 10.1073/pnas.0511311103

Michalska, K. J., Decety, J., Zeffiro, T. A., & Lahey, B. B. (2015). Association of regional gray matter volumes in the brain with disruptive behavior disorders in male and female children. *NeuroImage: Clinical*, 7, 252-257. doi: 10.1016/j.nicl.2014.12.012

Michalska, K. J., Zeffiro, T. A., & Decety, J. (2016). Brain response to viewing others being harmed in children with conduct disorder symptoms. *Journal of child psychology and psychiatry*, 57(4), 510-519. doi: 10.1111/jcpp.12474

Moffitt, T. E. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. United Kingdom: Cambridge University Press.

Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage*, 16(3), 696-703. doi: 10.1006/nimg.2002.1118

Nakai, T., Muraki, S., Bagarinao, E., Miki, Y., Takehara, Y., Matsuo, K., . . . Isoda, H. (2004). Application of independent component analysis to magnetic resonance imaging for enhancing the contrast of gray and white matter. *Neuroimage*, 21(1), 251-260. doi: 10.1016/j.neuroimage.2003.08.036

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 Research: Neuroimaging*, 98(1), 1-13. doi: 10.1016/S0925-4927(99)00044-X

Olson, I. R., Plotzker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*, 130(7), 1718-1731. doi: 10.1093/brain/awm052

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

26

- Raine, A., Yang, Y., Narr, K. L., & Toga, A. W. (2011). Sex differences in orbitofrontal gray as a partial explanation for sex differences in antisocial personality. *Molecular psychiatry*, 16(2), 227. doi: 10.1038/mp.2009.136
- Raschle, N. M., Menks, W. M., Fehlbauer, L. V., Steppan, M., Smaragdi, A., Gonzalez-Madruga, K., ... & Bernhard, A. (2018). Callous-unemotional traits and brain structure: Sex-specific effects in anterior insula of typically-developing youths. *NeuroImage: Clinical*, 17, 856-864. doi: 10.1016/j.nicl.2017.12.015
- Rosenblatt, J. D. (2016). Multivariate revisit to "sex beyond the genitalia". *Proceedings of the National Academy of Sciences of the United States of America*, 113(14), E1966.
- Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1995). Sex differences in crime: Do means and within-sex variation have similar causes? *Journal of Research in Crime and Delinquency*, 32(1), 84-100. doi: 10.1177/0022427895032001004
- Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M.-C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews*, 39, 34-50. doi: 10.1016/j.neubiorev.2013.12.004
- Rushton, J. P. (1992). Cranial capacity related to sex, rank, and race in a stratified random sample of 6,325 US military personnel. *Intelligence*, 16(3), 401-413. doi: 10.1016/0160-2896(92)90017-L
- Russett, C. E. (2009). *Sexual Science*. Cambridge, Massachusetts: Harvard University Press.
- Sacher, J., Neumann, J., Okon-Singer, H., Gotowiec, S., & Villringer, A. (2013). Sexual dimorphism in the human brain: evidence from neuroimaging. *Magnetic resonance imaging*, 31(3), 366-375. doi: 10.1016/j.mri.2012.06.007
- Sharpe, G., & Gelsthorpe, L. (2009). Engendering the Agenda: Girls, Young Women and Youth Justice. *Youth Justice*, 9(3), 195-208. doi: 10.1177/1473225409345098

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

- Steele, V. R., Claus, E. D., Aharoni, E., Vincent, G. M., Calhoun, V. D., & Kiehl, K. A. (2015). Multimodal imaging measures predict rearrest. *Frontiers in Human Neuroscience*, 9, 425. doi: 10.3389/fnhum.2015.00425
- Steele, V. R., Rao, V., Calhoun, V. D., & Kiehl, K. A. (2015). Machine learning of structural magnetic resonance imaging predicts psychopathic traits in adolescent offenders. *NeuroImage*, 145 (b), 265-273. doi: 10.1016/j.neuroimage.2015.12.013
- Steffensmeier, D., & Allan, E. (1996). Gender and crime: Toward a gendered theory of female offending. *Annual Review of Sociology*, 22(1), 459-487. doi: 10.1146/annurev.soc.22.1.459
- Tiihonen, J., Rossi, R., Laakso, M. P., Hodgins, S., Testa, C., Perez, J., ... & Könönen, M. (2008). Brain anatomy of persistent violent offenders: more rather than less. *Psychiatry Research: Neuroimaging*, 163(3), 201-212. doi: 10.1016/j.psychres.2007.08.012
- Turner, J. A., Calhoun, V. D., Michael, A., Van Erp, T. G., Ehrlich, S., Segall, J. M., ... & Bustillo, J. (2012). Heritability of multivariate gray matter measures in schizophrenia. *Twin Research and Human Genetics*, 15(3), 324-335. doi: 10.1017/thg.2012.1
- Vapnik, V. N. (1998). *Statistical Learning Theory* (Vol. 1). New York, NY: Wiley.
- Wilensky, A.E., Schafe, G.E., Kristensen, M.P., LeDoux, J.E. (2006). Rethinking the fear circuit: the central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. *Journal of Neuroscience*, 26, 12387-12396. doi: 10.1523/JNEUROSCI.4316-06.2006
- Xu, L., Groth, K. M., Pearlson, G., Schretlen, D. J., & Calhoun, V. D. (2009). Source-based morphometry: The use of independent component analysis to identify gray matter differences with application to schizophrenia. *Human Brain Mapping*, 30(3), 711-724. doi: 10.1002/hbm.20540

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

28

Yaesoubi, M., Miller, R. L., & Calhoun, V. D. (2015). Mutually temporally independent connectivity patterns: A new framework to study the dynamics of brain connectivity at rest with application to explain group difference based on gender. *Neuroimage*, 107, 85-94. doi: 10.1016/j.neuroimage.2014.11.054

For Peer Review

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Table 1. Sample Characteristics

	Full Sample (N = 1300)	Males only (n = 1014)	Females only (n = 286)	<u>t-value</u>	<u>p-value</u>
Age	31.3, 10.8	31.1, 11.1	32.0, 9.8	1.35	.178
IQ	95.8, 13.1	96.1, 13.7	95.0, 10.7	1.34	.181

*Note: Sample contains right handed individuals only; t and p values correspond to comparisons demonstrating no significant differences between males and females in age and IQ.*

Table 2.  
SVM Classification of Male and Female Offenders (Training Sample, N = 930)

	Classified as Female	Classified as Male
Females n = 211	201	10
Males n = 719	3	716
	204 Classified as Female	726 Classified as Male

Table 3.  
Logistic Regression Classification of Male and Female Offenders (Training Sample, N = 930)

	Classified as Female	Classified as Male
Females n = 211	191	20
Males n = 719	11	708
	202 Classified as Female	728 Classified as Male

Table 4.  
SVM Classification of Male and Female Offenders (Testing Sample, N = 370)

	Classified as Female	Classified as Male
Females n = 75	66	9
Males n = 295	14	281
	80 Classified as Female	290 Classified as Male

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Table 5.

Logistic Regression Classification of Male and Female Offenders (Testing Sample, N = 370)

	Classified as Female	Classified as Male
Females n = 75	62	13
Males n = 295	13	282
	75 Classified as Female	295 Classified as Male

Table 6.

SVM Classification of Non-Incarcerated Males and Females (Testing Sample, N = 526)

	Classified as Female	Classified as Male
Females n = 296	286	10
Males n = 230	24	206
	310 Classified as Female	216 Classified as Male

Table 7.

Logistic Regression Classification of Non-Incarcerated Males and Females (Testing Sample, N = 526)

	Classified as Female	Classified as Male
Females n = 296	284	12
Males n = 230	19	211
	303 Classified as Female	223 Classified as Male

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

**Figure Legends**

**Fig. 1** Series represents a generic human brain, sliced top to bottom in 4mm slices. Colored regions represent the gray matter volume (modulated) associated with a specific component derived from Source Based Morphometry (SBM). Blue represents component areas with higher loading for males ( $t = 14.02, p < .001$ ). Orange represents component areas with higher loading for females ( $t = 10.77, p < .001$ ). Higher loadings are dimensionally proportional to higher volume in these regions, but do not represent direct differences between males and females (see Figure 3). Colored bars are z-values for the spatial extent of the individual components across subjects.

**Fig. 2** Series represents a generic human brain, sliced top to bottom in 4mm slices. Colored regions represent the gray matter density (unmodulated) associated with a specific component derived from Source Based Morphometry (SBM). Blue represents component areas with higher loading for males ( $t = 6.02, p < .001$ ). Orange represents component areas with higher loading for females ( $t = 4.53, p < .001$ ). Higher loadings are dimensionally proportional to higher gray matter concentration in these regions, but do not represent direct differences between males and females (see Figure 3). Colored bars are z-values for the spatial extent of the individual components across subjects.

**Fig.3** T-tests demonstrating volumetric comparisons between males and females in the incarcerated (A) and non-incarcerated (B: Genomics Superstruct) samples. Orange maps show gray matter relatively larger in females; Blue maps show gray matter relatively larger in males. These figures are intended as a succinct summary of regional gray matter differences for ease of viewing; however, classification models were based on 57 components of gray matter derived from source based morphometry (e.g. Figures 1 & 2).



## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

32

Figure 1.

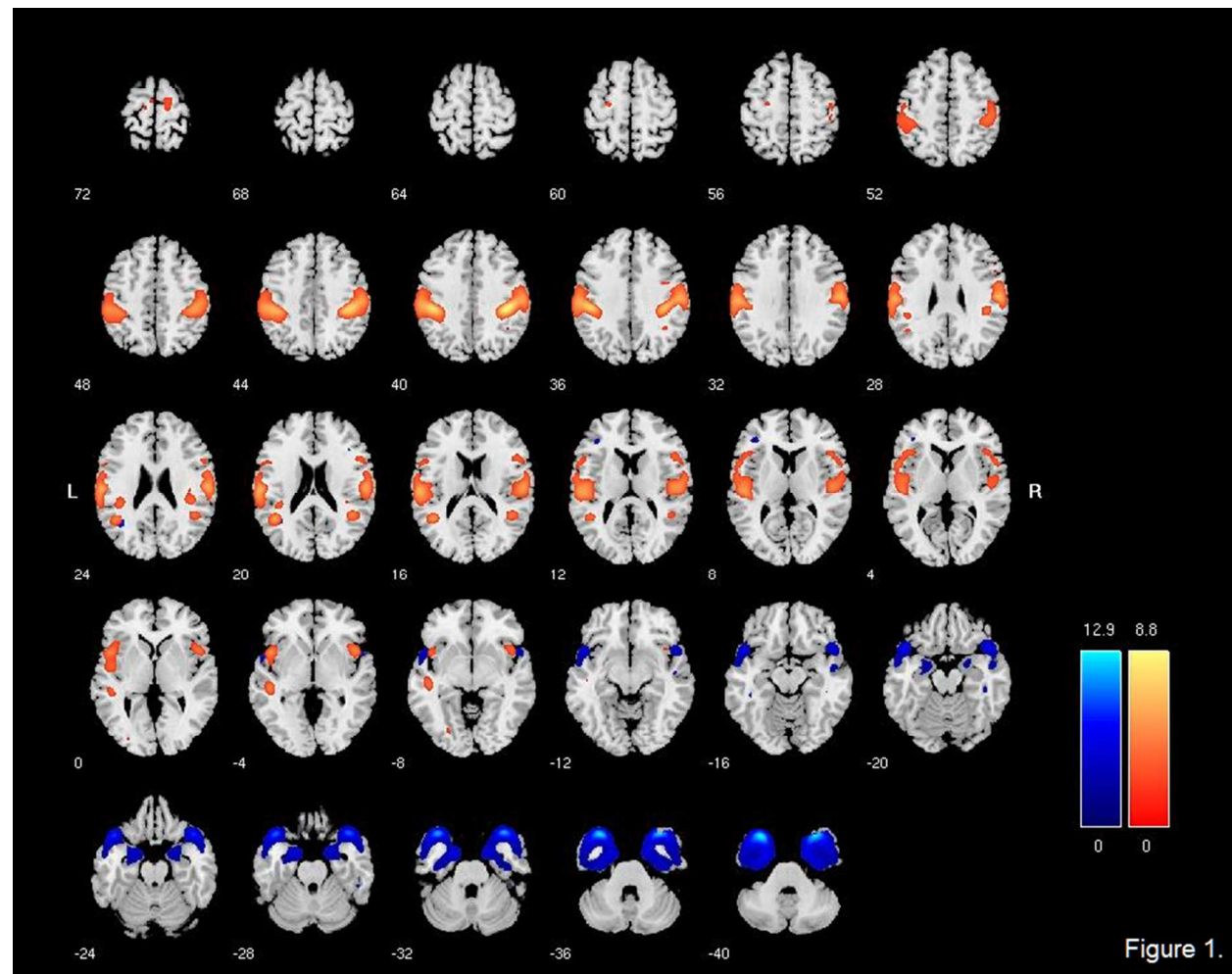


Figure 1.

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Figure 2.

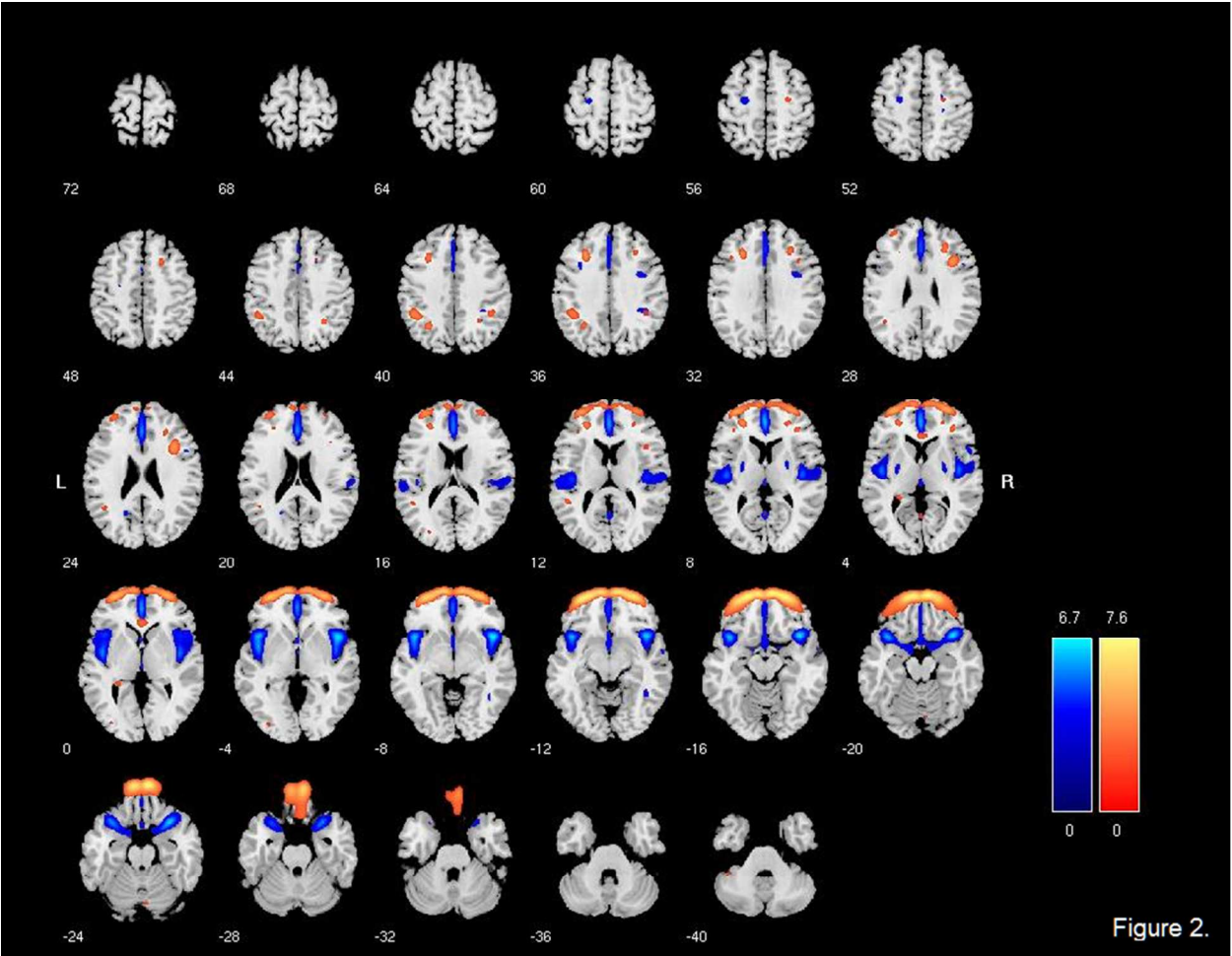
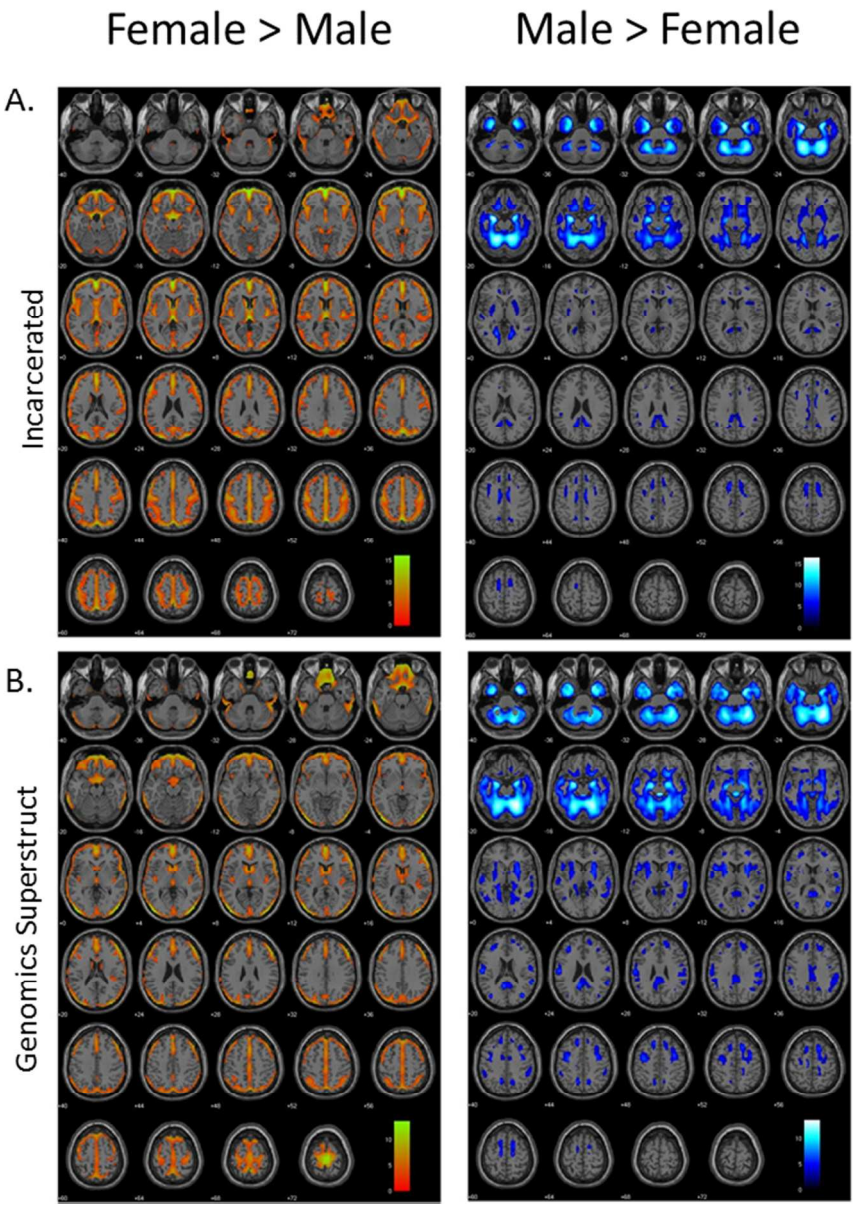


Figure 2.

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Figure 3.



**Machine Learning of Brain Gray Matter Differentiates Sex in a Large Forensic Sample**

Nathaniel E. Anderson <sup>a\*</sup>, Keith A. Harenski <sup>a</sup>, Carla L. Harenski <sup>a</sup>, Michael R. Koenigs <sup>b</sup>, Jean Decety <sup>c</sup>,  
Vince D. Calhoun <sup>a,d</sup>, Kent A. Kiehl <sup>a,d</sup>

- a. The Mind Research Network & Lovelace Biomedical and Environmental Research Institute,  
Albuquerque, NM, USA
- b. University of Wisconsin-Madison, Madison, WI, USA
- c. University of Chicago, Chicago, IL, USA
- d. University of New Mexico, Albuquerque, NM, USA

**Acknowledgements:** This research was supported in part by grants from the National Institute of Mental Health: R01MH070539, R01MH071896, R01MH085010, R01DA026505, R01DA026964, R01DA020870, R01EB005846, P20GM103472, and by The MacArthur Foundation Law and Neuroscience Project. We also thank the staff and inmates at New Mexico and Wisconsin Departments of Corrections; without their support, this research would not be possible.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

2

**Abstract**

Differences between males and females have been extensively documented in biological, psychological, and behavioral domains. Among these, sex differences in the rate and typology of antisocial behavior remains one of the most conspicuous and enduring patterns among humans. However, the nature and extent of sexual dimorphism in the brain among antisocial populations remains mostly unexplored. Here we seek to understand sex differences in brain structure between incarcerated males and females in a large sample ( $n=1300$ ) using machine learning. We apply source-based morphometry (SBM), a contemporary multivariate approach for quantifying gray matter measured with magnetic resonance imaging (MRI), and carry these parcellations forward using machine learning to classify sex. Models using components of brain gray matter volume and concentration were able to differentiate between males and females with greater than 93% generalizable accuracy. Highly differentiated components include orbitofrontal and frontopolar regions, proportionally larger in females, and anterior medial temporal regions proportionally larger in males. We also provide a complimentary analysis of a non-forensic healthy control sample and replicate our 93% sex discrimination. These findings demonstrate that the brains of males and females are highly distinguishable. Understanding sex differences in the brain has implications for elucidating variability in the incidence and progression of disease, psychopathology, and differences in psychological traits and behavior. The reliability of these differences confirms the importance of sex as a moderator of individual differences in brain structure and suggests future research should consider sex specific models.

**Keywords:** Antisocial Behavior, Sex, Gender, MRI, Source Based Morphometry, Machine Learning



CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Differences between males and females have been extensively documented in biological, psychological, and behavioral domains including criminology. These differences have encouraged the examination of sexual dimorphism and functional specialization in the human brain, which has remained a prominent branch of psychology and neuroscience research for well over a century (Broca, 1861; Lee & Pearson, 1901). Enduring controversy surrounds this topic as, historically, some have used the identification of organic differences among groups as a platform for reinforcing stereotypes and inferring a basis for fundamental limitations in ability, intelligence, and social status (Bean, 1906; Lynn, 1994; Russett, 2009). These findings and their interpretations have been rightfully challenged and debated for decades (Alper, 1985; Gould, 1996; Lynn, 1994). While developments in neuroimaging technology and computational methods have added new perspectives to the debate, they are no less prone to the controversies of *neurosexism* (Fine, 2013; Fine, Jordan-Young, Kaiser, & Rippon, 2013). Still, the elucidation of sex differences in the brain remains a highly important topic as these data help us understand more about the variable incidence and expression of mental health issues, disease, and our socio-biological and behavioral diversity as a species.

Despite the historically divisive nature of the topic, prevailing evidence at least does not favor a null hypothesis. The existence of organic differences in the brain, attributable to sex, has been well-replicated and reviewed at an aggregate level (Ruigrok et al., 2014; Sacher, Neumann, Okon-Singer, Gotowiec, & Villringer, 2013). Evidence suggests that men have larger brains than women, both absolutely and after controlling for average differences in body and head size (Allen, Damasio, Grabowski, Bruss, & Zhang, 2003; Lüders, Steinmetz, & Jäncke, 2002; Nopoulos, Flaum, O’Leary, & Andreasen, 2000). Morphologically, men’s brains have higher relative proportions of white matter while women’s brains have higher proportions of gray matter (Allen et al., 2003; Gur et al., 1999). Localized dimorphic regions are more variable in the literature, but accumulating evidence suggests that males

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

exhibit proportionally larger volumes in subcortical and interior cortical regions including the amygdala, hippocampus, parahippocampal gyrus, posterior cingulate, and temporal poles. Females, by contrast, exhibit proportionally larger volumes in several frontal cortical regions including inferior frontal gyrus, middle frontal gyrus, frontal pole, and frontal operculum (Ruigrok et al., 2014; Sacher et al., 2013). Advances in the quantification of brain data have the potential to improve the sensitivity of these measures—for instance, the parcellation and boundaries of individual anatomical features often vary across studies, and methods that apply data driven approaches for defining these regions may provide additional benefits (Gupta et al., 2015; Xu, Groth, Pearlson, Schretlen, & Calhoun, 2009).

Recent descriptions have emphasized that despite the presence of well-replicated average differences in many brain regions between sexes, there is substantial overlap in the morphology of individual brain regions. Based on a sample of over 1400 individuals, Joel and colleagues (2014) argued that a lack of internal consistency in dimorphism on many individual metrics makes male and female brains essentially indistinguishable at an individual level. Critics, however, responded that these conclusions rely on univariate logic (Del Giudice et al., 2015, 2016; Rosenblatt, 2016), and others have demonstrated that multivariate classification techniques relying on variation among many brain regions simultaneously are quite reliable. For example, relying on a sample of over 1500 individuals, Chekroud, Ward, Rosenberg, & Holmes (2016) achieved multivariate classification of sex based on gray matter with over 90% accuracy (see also Lao et al., 2004).

**Sex Differences and Antisocial Behavior**

Sex differences in the brain also covary with many observable differences in behavior. A largely unexplored segment of this research lies in extending this work to better understand antisocial behavior in forensic samples. Conspicuous sex differences are apparent in both the frequency and nature of antisocial behavior (Heidensohn & Silvestri, 2012; Moffitt, 2001). Men engage in antisocial behavior more frequently than females (Del Giudice, 2015; Rowe, Vazsonyi, & Flannery, 1995), and these

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

differences are even more pronounced when focusing on aggressive behavior and violent crime (Archer, 2004). For example, approximately 75% of those arrested in the United States are males, and 80% of those arrested for violent crimes are males (Federal Bureau of Investigation, 2014). A number of theories examine different influences on these persistent trends from the perspectives of sociology (Harris, 2000; Steffensmeier & Allan, 1996), psychology (Bettencourt & Miller, 1996; Del Giudice, 2015; Feder, Levant, & Dean, 2007), and biology (Batrinos, 2012; Ferguson & Beaver, 2009). Neuroscience may have a unique place among these perspectives if we consider that, regardless of the variety of antecedents, and their interactions, these influences likely converge in the brain as it is the organ that governs our behavior most proximally (see also Jorgensen, Anderson, & Barnes, 2016). For example, Raine, Yang, Narr, and Toga (2011) reported that reductions in orbitofrontal and middle frontal gray matter volume accounted for approximately 77% of the difference in antisocial behavior between males and females. Furthermore, relationships between brain variables and antisocial behavior differentiated by sex emerge quite early in development, with boys and girls showing unique associations between brain measures and delinquent conduct (Michalska, Decety, Zeffiro, & Lahey, 2015; Michalska, Zeffiro, & Decety, 2016; Raschle et al., 2018;). In short, sex differences in the brain appear to be important for understanding differences in antisocial behavior; however, it has yet to be demonstrated whether such differences amount to simple average differences or more precisely represent complex dimorphism between sexes in an antisocial sample.

Here we set out to examine the utility of morphological gray matter differences for differentiating between male and female incarcerated offenders. We further aim to demonstrate the utility of an advanced multivariate approach for quantifying gray matter (source-based morphometry) and to apply machine learning algorithms to these data to examine their separability based on sex. These data will be influential for understanding the nature of sex differences in the brain and a step toward further clarifying their role in large-scale behavioral patterns in our society.



## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

6

**Methods****Participants**

Data were collected from adult volunteers incarcerated in prisons in New Mexico and Wisconsin, and included male and female adult and juvenile offenders. A total of 1300 participants (males  $n = 1014$ ; females  $n = 286$ ) were included in this study. This sample was restricted to right-handed individuals. Ages ranged from 12 – 66 ( $M = 31.3$ ,  $SD = 10.8$ ). IQ ranged from 71 – 140 ( $M = 95.8$ ,  $SD = 13.1$ ); volunteers with  $IQ < 70$  were excluded from participation. Males and females did not differ on age or IQ, see sample characteristics in Table 1. Age and IQ were both used as covariates in prediction models. To observe best practices in prediction modeling approaches, this sample was randomly divided into two separate samples: one designated for training the machine ( $n = 930$ ), and the other sample dedicated for testing generalizability of the model ( $n = 370$ ), see classification methods below.

These data have been aggregated from participants who have volunteered for one of several ongoing brain imaging studies taking place in our collaborating forensic institutions. All research protocols have been approved by institutional review boards at the Ethical and Independent Review Services (E&I IRB) and University of Wisconsin-Madison and the Office of Human Research Protections (OHRP). This research meets ethical standards for responsible research with human subjects. Volunteers from forensic facilities were recruited via fliers and word of mouth. Meetings were scheduled with interested participants, and informed consent was obtained from all adult participants included in the study. Juvenile participants provided written informed assent in addition to their parent/guardian's written informed consent. Participants were informed of their right to terminate participation at any point, the lack of any direct institutional benefits, and that their participation would not affect their facility status or parole status. They were compensated with an hourly rate commensurate with standard rate for work assignments at their facility. Participants were screened for MRI safety, and

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

excluded for any contraindications (e.g. metal in body). Participants were also excluded from our studies if they reported a history of major head injury.

**MRI Data Acquisition and Analysis**

High-resolution T1-weighted structural MRI scans were acquired using the Mind Research Network Mobile Siemens 1.5T Avanto MRI scanner. This custom unit is specifically designed for transport and deployment within the secure environments of forensic facilities. A multi-echo MPAGE pulse sequence (repetition time=2530ms, echo times=1.64ms, 3.50ms, 5.36ms, 7.22ms, inversion time=1100ms, flip angle=7°, slice thickness=1.3mm, matrix size=256×256) was used, yielding 128 sagittal slices with an in-plane resolution of 1.0mm×1.0mm. Data were pre-processed and analyzed using Statistical Parametric Mapping software (SPM12; <http://www.fil.ion.ucl.ac.uk/spm>). T1 images were manually inspected by an operator blind to subject identity and realigned to ensure proper spatial normalization. Images were then analyzed via the unified segmentation approach as implemented in SPM12 (Ashburner & Friston, 2005). Unified segmentation allows for image registration based on Gaussian mixture modelling, tissue classification with warped prior probability maps and bias correction to be combined in the same generative model. Both volume (modulated data) and density (unmodulated data) were extracted for analyses. A Jacobian modulation was performed to preserve total volume (Ashburner & Friston, 2000; 2005). A nonlinear transformation without Jacobian determinants was performed on unmodulated images to extract grey matter density (Ashburner & Friston, 2000; 2005). Modulated and unmodulated images were resampled to 2x2x2 mm and smoothed with a 10mm full-width at half-maximum (FWHM) Gaussian kernel. Voxels with grey matter value of <0.15 were excluded in order to remove possible edge effects between grey matter and white matter. Only grey matter segments were used in this analysis.

Source Based Morphometry (SBM) was used to separate gray matter volume and density into maximally independent *source networks*. SBM is a multivariate alternative to voxel-based morphometry

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

(VBM). It is an application of Independent Component Analysis (ICA) for structural brain data (Calhoun & Adali, 2006; Xu et al., 2009). In contrast to VBM, SBM uses relationships among voxels, identifying distinct regions with common covariation between subjects. These source networks are spatially distinct and maximally independent (see also Caprihan et al., 2011; Nakai et al., 2004). SBM was carried out using spatial ICA as implemented in the GIFT Toolbox (Calhoun, Liu, & Adali, 2009) (<http://mialab.mrn.org/software/gift>). The number of source networks for modulated and unmodulated gray matter was set at 30 each, which has been demonstrated to produce a stable set of structural networks (Turner et al., 2012; Xu et al., 2009). Final component selection included the removal of spatial maps that were identified as noise or motion related. One modulated and two unmodulated components were removed, yielding 57 independent components used in the final analysis. Loading coefficients for each source network for each individual are carried forward as features in the classification models described below.

**Machine Learning Classification**

Loading coefficients for modulated and unmodulated gray matter from SBM analysis were used together with age and IQ estimates to test their utility in classifying sex (as categorically male or female). A large number of data classification approaches are available, each with certain advantages. While the mathematical approaches differ, these machine learning methods serve essentially the same purpose: to use a set of known information to build a generalizable model capable of predicting unknown information from a set of cases with limited data. In order to test the utility of a given type of information (for our purposes, localized gray matter volumes and density) for differentiating among groups/classes (sex), we provide full information on a set of cases for deriving the initial model, then test the performance (generalizability) of that model on cases for which limited data are provided, but for which the true value of a predicted outcome is known, such that errors in classification can be detected (i.e. supervised learning). In order to examine classification performance under best practices

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

for reproducibility (Friedman, Hastie, & Tibshirani, 2001; Vapnik, 1998), we first separated our data into a training sample ( $n = 930$ ) and a testing sample ( $n = 370$ ), based on random computer-aided assignment. Individuals in the testing sample were not included in model development to avoid potential bias. Their data were solely used for the purpose of validating the generalizability of the classification performance of models developed on the training set.

To compare among a number of methods, we used a publicly available toolbox which runs multiple classification approaches simultaneously to explore their relative performance. The *Polyssifier* toolbox (<https://github.com/alvarouc/polyssifier>) compares a k-nearest neighbors, linear support vector machine, radial basis function (RBF) support vector machine (SVM), decision tree, random forest, logistic regression, naïve Bayes, and voting classifier approaches. For the present data, SVM and logistic regression approaches exhibited the highest accuracy in performance, as we have found in several other datasets examining classification using brain data from incarcerated samples (e.g. Cope et al., 2014; Kiehl et al., 2018; Steele et al. 2015; Steele, Rao, Calhoun, & Kiehl, 2017). Here we reserve our descriptions of methods and outcomes to RBF SVM and logistic regression. All classifications and model testing were run in the *R* software package using the CARET (for SVM) (Kuhn et al., 2016) and glmnet (for logistic regression) libraries (Friedman, Hastie, & Tibshirani, 2010).

In SVM classification (Friedman et al., 2001; Vapnik, 1998) the goal is to construct a hyperplane separating the training examples within the feature space into two subspaces. The classification procedure takes place in two stages: training and testing. During training the algorithm finds a hyperplane (i.e. a high dimensional plane) that maximizes the margin between the training samples of both classes. The training samples that lie on the margins of the hyperplane are the called support vectors. These are the units/participants most difficult to classify and therefore define the boundaries of the hyperplane. Kernel functions can be applied if the data is not linearly separable in the original space and allows for group classification based on non-linear effects, e.g. RBF SVM. Once the decision

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

10

boundary is learned from the training data, it can be used to predict the class of new test samples.

Subject classification and best parameter selection via grid search over a reasonable range of values was accomplished using five repetitions of 10-fold cross validation in *R*.

For comparison, we also implemented an *elastic net* penalized logistic regression model. This method is designed to classify data based on sets of highly correlated predictors. It combines the properties of LASSO and ridge regression for promoting automatic variable selection while reducing the impact of highly correlated predictors. This method is theoretically appropriate due to the correlated nature of brain data within individuals (Bunea et al., 2011). Both modulated and unmodulated data from gray matter were used for prediction as both measures may be incrementally useful for differentiation of individual differences (e.g. Meda et al. 2008). For both SVM and logistic regression classification methods, models were developed on the training sample ( $n=930$ ) and then applied, unchanged, to the hold-out testing sample ( $n=370$ ).

From 30 modulated and unmodulated components derived from SBM, 3 components (1 modulated, 2 unmodulated) were determined to be noise and/or motion-related, and these were removed from further analyses. Each classification method then utilized brain data from 57 (29 modulated, 28 unmodulated) components. Implementation of a feature selection step did not improve overall performance of these models. Likewise, simple exclusion of any individual components, list-wise, did not significantly reduce prediction accuracy, suggesting all components are simultaneously, and incrementally useful in the resulting models. In other words, no single component can be meaningfully identified as the most important to classification, apart from other features of the model. In order to select and illustrate component maps (brain areas) contributing to classifications, we performed independent samples t-tests comparing males and females using the loading coefficients on each of the 57 components from SBM. Modulated and unmodulated components showing the highest *t*-values for *males > females* and *females > males* are presented for illustration purposes; however, their

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

relative value in prediction models requires the simultaneous inclusion of all components utilized in the final models. Individuals’ SVM loadings are dimensional and scale proportionally with gray matter volume/density, but loadings are not directly observable as discrete differences between groups.

To illustrate these differences, we have also carried out two-sample t-tests on gray matter volume (modulated VBM data) for the full sample of incarcerated males and females (n = 1300), and a non-incarcerated comparison group. This provides a more direct visualization of discrete gray matter differences between males and females.

**Non-Incarcerated Comparison Sample**

In order to compare and demonstrate the utility of the described SBM method for classifying sex in a non-incarcerated sample, we applied the same methods described above to the publicly available dataset from the Brain Genomics Superstruct Project ([www.neuroinfo.org/gsp/](http://www.neuroinfo.org/gsp/)). We restricted this sample to right-handed individuals only, yielding a total of n = 1448 (male/female). This sample was split randomly into training (n = 922; 550 female) and testing (n = 526; 296 female) sets. A total of 30 modulated and 30 unmodulated gray matter components were defined via SBM. Two modulated and two unmodulated components were determined to be noise yielding a total of 56 total components (28 modulated, 28 unmodulated) carried into the final prediction models. Mirroring analyses above, SVM and penalized logistic regression were carried out to classify males and females.

**Results**

Classification tables (training and testing samples) are provided for both SVM and elastic net logistic regression (see Tables 2-5). Cross-validated training samples yielded overall classification accuracy at 98.6% and 96.7% for SVM and logistic regression, respectively. Testing validation on the naïve hold-out set revealed generalizable overall classification accuracy at 93.8% for SVM and 93.0% for logistic regression.

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

12

In order to illustrate the structural components that are used as features in these classifications, independent samples *t*-tests comparing males and females on individuals' loading coefficients in SPSS v20 were carried out on all 57 (29 modulated, 28 unmodulated) SBM components. A selection of components demonstrating significant differences between males and females is provided in Figures 1 & 2. These component maps show the spatial organization of a selection of four (out of 57 total) structural networks that were useful in the classification models. These four were selected by carrying out *t*-tests between males and females in the training sample ( $n = 930$ ) on the loading coefficients for all components. For illustrative purposes components in loadings for gray matter volume (modulated; Figure 1) and concentration (unmodulated; Figure 2) were larger for males (blue) and females (orange) are presented. These maps do not represent volumetric differences, per se, between men and women, but simply the organization of structural components across the entire sample. Increasing volume or density in the highlighted regions is distinguishing of males and females in the classification of this sample. A component that is significantly larger for females, for example indicates that females have larger loading coefficients for that component than males, and these differences (across 57 components) were ultimately useful for classification. These loading coefficients thus scale proportionally with volumetric differences, which can be directly visualized with two-sample *t*-tests of VBM data.

**VBM Comparison**

*T*-tests directly comparing gray matter volume and density between males and females are shown in Figure 3. These brain maps summarize voxel-by-voxel statistics across the entire brain illustrating areas with relatively higher gray matter volume in males and females. SBM component differences and *t*-test comparisons confirm prominent regional differences include orbitofrontal and frontopolar regions, proportionally larger in females, and anterior/medial temporal regions, proportionally larger in males (see Figure 3).

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

**Non-Incarcerated Comparison Sample**

SBM-based classification results for the non-incarcerated comparison sample were virtually identical to those from the incarcerated sample. Classification on the naïve testing data (n = 526) produced overall accuracy of approximately 93.5% for SVM and 94.1% for penalized logistic regression methods. Confusion matrices for these methods are provided in Tables 6 & 7.

**Discussion**

The purpose of this study was twofold. First, we set out to examine the reliability of sexual dimorphism in the brain among a large incarcerated sample, extending work carried out by others in non-incarcerated healthy control subjects (c.f. Chekroud et al., 2016). Second, we aimed to demonstrate the utility of an ICA-based quantification method of gray matter for predicting sex from brain data, rather than relying on pre-established anatomical divisions of the brain. As expected, our results demonstrate that sexual dimorphism in brain structure is highly apparent among incarcerated samples, and the multivariate methods used to quantify gray matter allowed greater than 93% accuracy in classifying individuals as male or female. These findings are an important incremental step for elucidating neurobiological influences on male and female criminal behavior, underscoring the importance of sex differences in antisocial brains. Further, this is the first demonstration of brain-based sex classifications in an incarcerated sample, demonstrating that multivariate patterns of gray matter are reliably dimorphic between sexes among male and female inmates. Finally, to our knowledge this is the first report to demonstrate that ICA-based parcellation of brain anatomy is as effective, if not more reliable, than traditional atlas-based approaches or voxel-based morphometry for defining sex-specific differences.

The implications of this work are multifaceted and stand to encourage more precise scientific work aimed at understanding of the biological influences on crime. Our data suggest that the brain



## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

14

structure of antisocial males and females are no less distinguishable than what has been demonstrated among normative samples. This supports the notion that antisocial behavior in women should also be examined independently from conclusions that have been drawn from a traditionally male-focused literature. It also emphasizes the importance of considering biological-behavioral patterns separately in each gender, or at least considering the moderating effects of sex in these outcomes.

More generally, our findings challenge recent suggestions that male and female brains are essentially indistinguishable (Joel et al., 2015). The present findings add to a number of demonstrations that multivariate patterns in brain anatomy are quite reliable at classifying individual male and female brains (Chekroud et al., 2016; Del Giudice et al., 2015, 2016; Lao et al., 2004). While we agree that considerable overlap exists in the distribution of individual brain measures, concluding their effective equivalence misrepresents the complexity of the issue. The consequence of this would be to mistakenly undermine the value of drawing on sex differences to better understand individual differences in healthy and disordered brain function and behavior, including important differences in the incidence and progression of disease and psychopathology.

We suspect that conclusions endorsing the uniformity of male and female brains are at least partially undergirded by a benevolent motivation to subvert sexist attitudes. As noted above, neuroanatomical sexual dimorphism is a perpetually controversial topic since it has often been used as a defense of prejudiced beliefs, sex-based stereotypes, and categorical limitations in abilities. Further, it has often been misappropriated to justify unfair or unequal treatment in social domains. What should be clear, however, is that sexist interpretations as well as the notion of equivalence between sexes both misconstrue and undermine the true value of understanding sex differences. The implicit presumption in both of these positions seems to be that *different* must mean *unequal*, branding one as *inferior*, but males and females do not need to be identical to be treated equally. Further, developing a better understanding of the origins and consequences of those differences may be imperative to continued

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

progress in behavioral neuroscience. Understanding the extent and boundaries of sex-based differences in the brain will ultimately help us better interpret differences in mental health issues, behavior, mortality, and the fundamental causes and contextual limitations of these differences.

The other important aim of this study was to demonstrate that a multivariate, data-driven technique for parceling independent components of brain anatomy (ICA) is an effective method for quantifying these sex-based differences. While sexual dimorphism in the brain has been studied for over a century, we are no longer limited to weighing post-mortem brains, as Broca did in the late 19<sup>th</sup> century (Broca, 1861), or inferring in vivo brain size from cranial circumference (Lee & Pearson, 1901; Rushton, 1992). Major historical innovations including magnetic resonance imaging have been coupled with increasingly sensitive analytic methods, like those applying ICA to MRI data. Improving the techniques and testing novel ways of quantifying brain data are instrumental steps in improving the utility of brain data for practical purposes. Comparing our data with studies that have used automatic parcellation, we demonstrate comparable, and even slightly improved classification accuracy. This may be the result of the ICA methods partitioning variance in a way that is optimized for quantifying individual differences in a way that increases sensitivity (Allen, Erhardt, Wei, Eichele, & Calhoun, 2012; Calhoun, Pearlson, & Pekar, 2001).

Components derived from ICA demonstrate voxels of gray matter that vary systematically with one another across subjects. The loading coefficients from the ICA solution are what successfully classify individuals as males and females in this demonstration. These loading coefficients indicate the relative contribution of an individual to the ICA solution that gave rise to that component. Thus, individuals with smaller absolute loading coefficients for a given component are less important for identifying and quantifying that network. The illustrative components provided in Figures 1 and 2 should not be interpreted directly as volumetric differences between males and females. Instead, values in each component are indicative of a pattern of gray matter structure that is more discriminative

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

16

of males (blue) and females (orange) in our sample. Still, these loadings will scale proportionally with gray matter volume, and t-tests comparing volume directly across sex are provided for ease of viewing.

Anatomically, Figure 1 shows a component of increasing gray matter volume around the amygdala and anterior temporal cortex that is discriminating of males (blue area), while increased temporoparietal volume (BA40) extending into somatosensory areas is more discriminating of females (orange areas). Figure 2 shows that increased density in the anterior insula and medial prefrontal cortex is discriminating of males, while increasing density in the frontopolar and ventromedial/orbitofrontal gray matter is more discriminating of females. T-test maps (Figure 3) illustrate how these regions are systematically reflected in volumetric comparisons of males and females as well.

Importantly, brain regions differing between males and females have conspicuous roles in many cognitive features that align with previously identified patterns relevant to antisocial behavior. The amygdala plays a well-known role in detecting threat and in fear conditioning and coordination of responses to threat (Wilensky, Schafe, Kristensen, & LeDoux, 2006). Exaggerated activity in the amygdala and diminished orbitofrontal activity has been associated with patterns of reactive aggression and violence (Coccaro, McCloskey, Fitzgerald, & Phan, 2007; Meyer-Lindenberg et al., 2006). The anterior temporal cortex (temporal pole, BA38) is tightly connected to limbic and paralimbic structures influencing complex social and emotional processing (Olson, Plotzker, & Ezzyat, 2007). Lower volume and functioning in these regions have also been associated with diminished inhibition of violent and aggressive behavior (Blair, 2016; Coccaro et al., 2007; Tiihonen et al, 2008). The frontopolar and orbitofrontal regions are important in moral judgment and planning/executing behavior, including incorporating contingencies that involve consequences and other learned reinforcement (Greene, & Haidt, 2002; Moll, Oliveira-Souza, Bramati, & Grafman, 2002). The temporoparietal junction is also fundamental for execution of attentional shifts required for perspective-taking, theory of mind, and empathy (Decety & Lamm, 2007). It is noteworthy, however, that these prevailing patterns in antisocial

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

and externalizing populations have been studied almost exclusively among males. Still, it remains reasonable to suspect that higher relative rates of antisocial behavior, and particularly violent-antisocial behavior, in males may be tied to sex-specific differences in these brain areas (cf. Raine et al. 2011).

While on-average differences (cf. Figure 3) in these regions have been recognized previously in healthy samples (Ruigrok et al., 2014; Sacher et al., 2013), it is conceptually important that the current work has demonstrated multivariate models that approach something closer to truly dimorphic patterns (see McCarthy et al. 2012), effectively differentiating individuals into categorical groups based on intrinsic structural networks (via SBM components). Demonstrating these patterns highlights that antisocial behavior in men and women occurs in the context of unique neural architecture that accompanies highly reliable dimorphic sex differences. It further underscores the importance of accounting for sex differences when conducting cognitive neuroscience research among forensic samples.

**Limitations, future directions, and conclusions**

The results of the present study are accompanied by a number of limitations that should promote additional research on these important topics. Our study examines a relatively simple model intended to account primarily for the influence of biological sex on volumetric brain data, but many other moderating variables and quantitative methods are likely to improve our understanding of this topic. The present descriptions of sexual dimorphism in the human brain remain relatively macroscopic and imprecise. Future studies should examine discriminability based on measures that include functional activity as well as structural and functional connectivity (e.g. Ingallhalikar et al., 2014; Yaesoubi et al., 2015). Sex-related differences in brain structure and function are likely to interact with age as well (Brain Development Cooperative Group, 2012; Lenroot et al., 2007). In this study, we used age as a covariate in our models, but future studies may examine how this discriminability changes more precisely as a function of developmental progress over the lifespan (see also Kiehl et al., 2018).

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

18

The present study examined sexual dimorphism among incarcerated offenders primarily to establish essential discriminability and underscore the value of considering the moderating effects of sex/gender in forensic neuroscience. We ardently condemn the historically slippery perspective of relying on established brain differences to justify inequitable treatment based on prejudice. We simultaneously caution against the conclusion that male and female brains are indistinguishable, as this has the potential to undermine the real utility of these data in understanding our diversity as a species. It should be clear from these data that males and females exhibit morphometric differences that are reliable enough to predict sex with high individual specificity. The way we choose to utilize this information can either improve ongoing scientific rigor or alternatively reinforce obsolete ideas and stagnate progress.

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

References

Adler, F. S., & Adler, H. M. (1975). *Sisters in crime: The rise of the new female criminal*. New York: McGraw-Hill.

Allen, E. A., Erhardt, E. B., Wei, Y., Eichele, T., & Calhoun, V. D. (2012). Capturing inter-subject variability with group independent component analysis of fMRI data: a simulation study. *Neuroimage*, 59(4), 4141-4159. doi: 10.1016/j.neuroimage.2011.10.010

Allen, J. S., Damasio, H., Grabowski, T. J., Bruss, J., & Zhang, W. (2003). Sexual dimorphism and asymmetries in the gray–white composition of the human cerebrum. *Neuroimage*, 18(4), 880-894. doi: 10.1016/S1053-8119(03)00034-X

Alper, J. S. (1985). Sex differences in brain asymmetry: A critical analysis. *Feminist Studies*, 11(1), 7-37. doi: 10.2307/3180130

Archer, J. (2004). Sex differences in aggression in real-world settings: a meta-analytic review. *Review of General Psychology*, 8(4), 291-322. doi: 10.1037/1089-2680.8.4.291

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, 9(3), 212-230. doi: 10.1207/s15327957pspr0903\_2

Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry--The Methods. *Neuroimage*, 11(6), 805-821. doi: 10.1006/nimg.2000.0582

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage*, 26(3), 839-851. doi: 10.1016/j.neuroimage.2005.02.018

Batrinis, M. L. (2012). Testosterone and aggressive behavior in man. *International journal of endocrinology and metabolism*, 10(3), 563-568. doi: 10.5812/ijem.3661

Bean, R. B. (1906). Some racial peculiarities of the Negro brain. *Developmental Dynamics*, 5(4), 353-432. doi: 10.1002/aja.1000050402

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

20

- Bettencourt, B., & Miller, N. (1996). Gender differences in aggression as a function of provocation: a meta-analysis. *Psychological Bulletin*, 119(3), 422-447. doi: 10.1037/0033-2909.119.3.422
- Blair, R. J. (2016). The neurobiology of impulsive aggression. *Journal of child and adolescent psychopharmacology*, 26(1), 4-9. doi: 10.1089/cap.2015.0088
- 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. *Cerebral Cortex*, 22(1), 1-12. doi: 10.1093/cercor/bhr018
- Broca, P. (1861). *Sur le volume et la forme du cerveau suivant les individus et suivant les races* (Vol. 1). Harvard University: Hennuyer.
- Bunea, F., She, Y., Ombao, H., Gongvatana, A., Devlin, K., & Cohen, R. (2011). Penalized least squares regression methods and applications to neuroimaging. *Neuroimage*, 55(4), 1519-1527. doi: 10.1016/j.neuroimage.2010.12.028
- Calhoun, V. D., & Adali, T. (2006). Unmixing fMRI with independent component analysis. *IEEE Engineering in Medicine and Biology Magazine*, 25(2), 79-90. doi: 10.1109/MEMB.2006.1607672
- Calhoun, V. D., Liu, J., & Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage*, 45(1), S163-S172. doi: 10.1016/j.neuroimage.2008.10.057
- Calhoun, V. D., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3), 140-151. doi: 10.1002/hbm.1048
- Campbell, A. (1981). *Girl Delinquents*. UK: Oxford: Basil Blackwell.
- Caprihan, A., Abbott, C., Yamamoto, J., Pearlson, G., Perrone-Bizzozero, N., Sui, J., & Calhoun, V. D. (2011). Source-based morphometry analysis of group differences in fractional anisotropy in schizophrenia. *Brain Connectivity*, 1(2), 133-145. doi: 10.1089/brain.2011.0015

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Carrington, K. (2006). Does feminism spoil girls? Explanations for official rises in female delinquency.

*Australian & New Zealand Journal of Criminology*, 39(1), 34-53. doi: 10.1375/acri.39.1.34

Chekroud, A. M., Ward, E. J., Rosenberg, M. D., & Holmes, A. J. (2016). Patterns in the human brain

mosaic discriminate males from females. *Proceedings of the National Academy of Sciences*,

113(14), E1968-E1968. doi: 10.1073/pnas.1523888113

Chesney-Lind, M., Morash, M., & Stevens, T. (2008). Girls troubles, girls' delinquency, and gender

responsive programming: A review. *Australian & New Zealand Journal of Criminology*, 41(1),

162-189. doi: 10.1375/acri.41.1.162

Coccaro, E. F., McCloskey, M. S., Fitzgerald, D. A., & Phan, K. L. (2007). Amygdala and orbitofrontal

reactivity to social threat in individuals with impulsive aggression. *Biological psychiatry*, 62(2),

168-178. doi: 10.1016/j.biopsych.2006.08.024

Cope, L. M., Ermer, E., Gaudet, L. M., Steele, V. R., Eckhardt, A. L., Arbabshirani, M. R., ... & Kiehl, K. A.

(2014). Abnormal brain structure in youth who commit homicide. *Neuroimage: clinical*, 4, 800-

807. doi: 10.1016/j.nicl.2014.05.002

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how

low-level computational processes contribute to meta-cognition. *The Neuroscientist*, 13(6), 580-

593. doi: 10.1177/1073858407304654

Del Giudice, M. (2015). Gender differences in personality and social behavior. In J. D. Wright (Ed.),

*International Encyclopedia of the Social and Behavioral Sciences* (2nd ed., pp. 750-756). New

York, NY: Elsevier.

Del Giudice, M., Lipka, R. A., Puts, D. A., Bailey, D. H., Bailey, J. M., & Schmitt, D. P. (2015). Mosaic

Brains? A Methodological Critique of Joel et al. (2015). Retrieved from:

[http://cogprints.org/10046/1/Delgiudice\\_etal\\_critique\\_joel\\_2015.pdf](http://cogprints.org/10046/1/Delgiudice_etal_critique_joel_2015.pdf)



## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

22

- Del Giudice, M., Lippa, R. A., Puts, D. A., Bailey, D. H., Bailey, J. M., & Schmitt, D. P. (2016). Joel et al.'s method systematically fails to detect large, consistent sex differences. *Proceedings of the National Academy of Sciences*, 113(14), E1965-E1965. doi: 10.1073/pnas.1525534113
- Federal Bureau of Investigation (2014). *Crime in the United States*. from <https://ucr.fbi.gov/crime-in-the-u.s/2014/crime-in-the-u.s.-2014/persons-arrested/main>
- Feder, J., Levant, R. F., & Dean, J. (2007). Boys and violence: A gender-informed analysis. *Professional Psychology: Research and Practice*, 38(4), 385-391. doi: 10.1037/0735-7028.38.4.385
- Ferguson, C. J., & Beaver, K. M. (2009). Natural born killers: The genetic origins of extreme violence. *Aggression and Violent Behavior*, 14(5), 286-294. doi: 10.1016/j.avb.2009.03.005
- Fine, C. (2013). Is there neurosexism in functional neuroimaging investigations of sex differences? *Neuroethics*, 6(2), 369-409. doi: 10.1007/s12152-012-9169-1
- Fine, C., Jordan-Young, R., Kaiser, A., & Rippon, G. (2013). Plasticity, plasticity, plasticity... and the rigid problem of sex. *Trends in Cognitive Sciences*, 17(11), 550-551. doi: 10.1016/j.tics.2013.08.010
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The Elements of Statistical Learning* (Vol. 1). New York: Springer.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). GLMNET: Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1-22.
- Gould, S. J. (1996). *The Mismeasure of Man*. New York, NY: WW Norton & Company.
- Greenfeld, L. A., & Snell, T. L. (2000). *Women Offenders*. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Gupta, C. N., Calhoun, V. D., Rachakonda, S., Chen, J., Patel, V., Liu, J., ... & Buitelaar, J. (2014). Patterns of gray matter abnormalities in schizophrenia based on an international mega-analysis. *Schizophrenia bulletin*, 41(5), 1133-1142. doi: 10.1093/schbul/sbu177

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. *Journal of Neuroscience*, 19(10), 4065-4072.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work?. *Trends in cognitive sciences*, 6(12), 517-523. doi: 10.1016/S1364-6613(02)02011-9

Harris, A. P. (2000). Gender, violence, race, and criminal justice. *Stanford Law Review*, 777-807.

Heidensohn, F. (1989). Gender and Crime. In *Crime and Society: Sociology for a Changing World* (pp. 85-111). London: Palgrave, Springer. doi: 10.1007/978-1-349-19763-7\_5

Heidensohn, F., & Silvestri, M. (2012). Gender and Crime. In M. Maguire, R. Morgan & R. Reiner (Eds.), *The Oxford Handbook of Criminology*. Oxford, UK: Oxford University Press.

Heimer, K. (Producer). (2000). Changes in the gender gap in crime and women’s economic marginalization. *The Nature of Crime: Continuity and Change*. Retrieved from [https://www.ncjrs.gov/criminal\\_justice2000/vol\\_1/02i.pdf](https://www.ncjrs.gov/criminal_justice2000/vol_1/02i.pdf)

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., . . . Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences*, 111(2), 823-828. doi: 10.1073/pnas.1316909110

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., . . . Margulies, D. S. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences*, 112(50), 15468-15473. doi: 10.1073/pnas.1509654112

Jorgensen, C., Anderson, N. E., & Barnes, J. (2016). Bad Brains: Crime and Drug Abuse from a Neurocriminological Perspective. *American Journal of Criminal Justice*, 41(1), 47-69. doi: 10.1007/s12103-015-9328-0

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

24

- Kiehl, K. A., Anderson, N. E., Aharoni, E., Maurer, J. M., Harenski, K. A., Rao, V., ... & Kosson, D. (2018). Age of gray matters: Neuroprediction of recidivism. *NeuroImage: Clinical*, 19, 813-823. doi: 10.1016/j.nicl.2018.05.036
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., . . . Hunt, T. (Producer). (2016). CARET: Classification and Regression Training, R package version 6.0-73. Retrieved from <https://cran.r-project.org/web/packages/caret/caret.pdf>
- Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S. M., & Davatzikos, C. (2004). Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*, 21(1), 46-57. doi: 10.1016/j.neuroimage.2003.09.027
- Lee, A. E., & Pearson, K. (1901). Data for the Problem of Evolution in Man: A First Study of the Correlation of the Human Skull. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical, and Engineering Sciences*, 196, 225-264. doi: 10.1098/rsta.1901.0005
- Lenroot, R. K., Gogtay, N., Greenstein, D. K., Wells, E. M., Wallace, G. L., Clasen, L. S., . . . Evans, A. C. (2007). Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage*, 36(4), 1065-1073. doi: 10.1016/j.neuroimage.2007.03.053
- Lüders, E., Steinmetz, H., & Jäncke, L. (2002). Brain size and grey matter volume in the healthy human brain. *NeuroReport*, 13(17), 2371-2374. doi: 10.1097/00001756-200212030-00040
- Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences*, 17(2), 257-271. doi: 10.1016/0191-8869(94)90030-2
- 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. *Journal of Neuroscience*, 32(7), 2241-2247. doi: 10.1523/JNEUROSCI.5372-11.2012
- Meda, S. A., Giuliani, N. R., Calhoun, V. D., Jagannathan, K., Schretlen, D. J., Pulver, A., ... & Sharma, T. (2008). A large scale (N= 400) investigation of gray matter differences in schizophrenia using

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

optimized voxel-based morphometry. *Schizophrenia research*, 101(1-3), 95-105. doi: 10.1016/j.schres.2008.02.007

Meyer-Lindenberg, A., Buckholtz, J. W., Kolachana, B., Hariri, A. R., Pezawas, L., Blasi, G., ... & Egan, M. (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proceedings of the National Academy of Sciences*, 103(16), 6269-6274. doi: 10.1073/pnas.0511311103

Michalska, K. J., Decety, J., Zeffiro, T. A., & Lahey, B. B. (2015). Association of regional gray matter volumes in the brain with disruptive behavior disorders in male and female children. *NeuroImage: Clinical*, 7, 252-257. doi: 10.1016/j.nicl.2014.12.012

Michalska, K. J., Zeffiro, T. A., & Decety, J. (2016). Brain response to viewing others being harmed in children with conduct disorder symptoms. *Journal of child psychology and psychiatry*, 57(4), 510-519. doi: 10.1111/jcpp.12474

Moffitt, T. E. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. United Kingdom: Cambridge University Press.

Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage*, 16(3), 696-703. doi: 10.1006/nimg.2002.1118

Nakai, T., Muraki, S., Bagarinao, E., Miki, Y., Takehara, Y., Matsuo, K., . . . Isoda, H. (2004). Application of independent component analysis to magnetic resonance imaging for enhancing the contrast of gray and white matter. *Neuroimage*, 21(1), 251-260. doi: 10.1016/j.neuroimage.2003.08.036

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 Research: Neuroimaging*, 98(1), 1-13. doi: 10.1016/S0925-4927(99)00044-X

Olson, I. R., Plotzker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*, 130(7), 1718-1731. doi: 10.1093/brain/awm052

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

26

- Raine, A., Yang, Y., Narr, K. L., & Toga, A. W. (2011). Sex differences in orbitofrontal gray as a partial explanation for sex differences in antisocial personality. *Molecular psychiatry*, 16(2), 227. doi: 10.1038/mp.2009.136
- Raschle, N. M., Menks, W. M., Fehlbauer, L. V., Steppan, M., Smaragdi, A., Gonzalez-Madruga, K., ... & Bernhard, A. (2018). Callous-unemotional traits and brain structure: Sex-specific effects in anterior insula of typically-developing youths. *NeuroImage: Clinical*, 17, 856-864. doi: 10.1016/j.nicl.2017.12.015
- Rosenblatt, J. D. (2016). Multivariate revisit to "sex beyond the genitalia". *Proceedings of the National Academy of Sciences of the United States of America*, 113(14), E1966.
- Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1995). Sex differences in crime: Do means and within-sex variation have similar causes? *Journal of Research in Crime and Delinquency*, 32(1), 84-100. doi: 10.1177/0022427895032001004
- Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M.-C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews*, 39, 34-50. doi: 10.1016/j.neubiorev.2013.12.004
- Rushton, J. P. (1992). Cranial capacity related to sex, rank, and race in a stratified random sample of 6,325 US military personnel. *Intelligence*, 16(3), 401-413. doi: 10.1016/0160-2896(92)90017-L
- Russett, C. E. (2009). *Sexual Science*. Cambridge, Massachusetts: Harvard University Press.
- Sacher, J., Neumann, J., Okon-Singer, H., Gotowiec, S., & Villringer, A. (2013). Sexual dimorphism in the human brain: evidence from neuroimaging. *Magnetic resonance imaging*, 31(3), 366-375. doi: 10.1016/j.mri.2012.06.007
- Sharpe, G., & Gelsthorpe, L. (2009). Engendering the Agenda: Girls, Young Women and Youth Justice. *Youth Justice*, 9(3), 195-208. doi: 10.1177/1473225409345098

CLASSIFICATION OF SEX BY BRAIN STRUCTURE

Steele, V. R., Claus, E. D., Aharoni, E., Vincent, G. M., Calhoun, V. D., & Kiehl, K. A. (2015). Multimodal imaging measures predict rearrest. *Frontiers in Human Neuroscience*, 9, 425. doi: 10.3389/fnhum.2015.00425

Steele, V. R., Rao, V., Calhoun, V. D., & Kiehl, K. A. (2015). Machine learning of structural magnetic resonance imaging predicts psychopathic traits in adolescent offenders. *NeuroImage*, 145 (b), 265-273. doi: 10.1016/j.neuroimage.2015.12.013

Steffensmeier, D., & Allan, E. (1996). Gender and crime: Toward a gendered theory of female offending. *Annual Review of Sociology*, 22(1), 459-487. doi: 10.1146/annurev.soc.22.1.459

Tiihonen, J., Rossi, R., Laakso, M. P., Hodgins, S., Testa, C., Perez, J., ... & Könönen, M. (2008). Brain anatomy of persistent violent offenders: more rather than less. *Psychiatry Research: Neuroimaging*, 163(3), 201-212. doi: 10.1016/j.psychres.2007.08.012

Turner, J. A., Calhoun, V. D., Michael, A., Van Erp, T. G., Ehrlich, S., Segall, J. M., ... & Bustillo, J. (2012). Heritability of multivariate gray matter measures in schizophrenia. *Twin Research and Human Genetics*, 15(3), 324-335. doi: 10.1017/thg.2012.1

Vapnik, V. N. (1998). *Statistical Learning Theory* (Vol. 1). New York, NY: Wiley.

Wilensky, A.E., Schafe, G.E , Kristensen, M.P., LeDoux, J.E. (2006). Rethinking the fear circuit: the central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. *Journal of Neuroscience*, 26, 12387-12396. doi: 10.1523/JNEUROSCI.4316-06.2006

Xu, L., Groth, K. M., Pearlson, G., Schretlen, D. J., & Calhoun, V. D. (2009). Source-based morphometry: The use of independent component analysis to identify gray matter differences with application to schizophrenia. *Human Brain Mapping*, 30(3), 711-724. doi: 10.1002/hbm.20540

## CLASSIFICATION OF SEX BY BRAIN STRUCTURE

28

Yaesoubi, M., Miller, R. L., & Calhoun, V. D. (2015). Mutually temporally independent connectivity patterns: A new framework to study the dynamics of brain connectivity at rest with application to explain group difference based on gender. *Neuroimage*, 107, 85-94. doi: 10.1016/j.neuroimage.2014.11.054

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 1. Sample Characteristics

	Full Sample (N = 1300)	Males only (n = 1014)	Females only (n = 286)	<u>t-value</u>	<u>p-value</u>
Age	31.3, 10.8	31.1, 11.1	32.0, 9.8	1.35	.178
IQ	95.8, 13.1	96.1, 13.7	95.0, 10.7	1.34	.181

*Note: Sample contains right handed individuals only; t and p values correspond to comparisons demonstrating no significant differences between males and females in age and IQ.*

For Peer Review



Table 2.  
SVM Classification of Male and Female Offenders (Training Sample, N = 930)

	Classified as Female	Classified as Male
Females n = 211	201	10
Males n = 719	3	716
	204 Classified as Female	726 Classified as Male

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 3.  
Logistic Regression Classification of Male and Female Offenders (Training Sample, N = 930)

	Classified as Female	Classified as Male
Females n = 211	191	20
Males n = 719	11	708
	202 Classified as Female	728 Classified as Male

For Peer Review

Table 4.  
SVM Classification of Male and Female Offenders (Testing Sample, N = 370)

	Classified as Female	Classified as Male
Females n = 75	66	9
Males n = 295	14	281
	80 Classified as Female	290 Classified as Male

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 5.  
Logistic Regression Classification of Male and Female Offenders (Testing Sample, N = 370)

	Classified as Female	Classified as Male
Females n = 75	62	13
Males n = 295	13	282
	75 Classified as Female	295 Classified as Male

For Peer Review

Table 6.  
SVM Classification of Non-Incarcerated Males and Females (Testing Sample, N = 526)

	Classified as Female	Classified as Male
Females n = 296	286	10
Males n = 230	24	206
	310 Classified as Female	216 Classified as Male

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 7.  
Logistic Regression Classification of Non-Incarcerated Males and Females (Testing Sample, N = 526)

	Classified as Female	Classified as Male
Females n = 296	284	12
Males n = 230	19	211
	303 Classified as Female	223 Classified as Male

For Peer Review

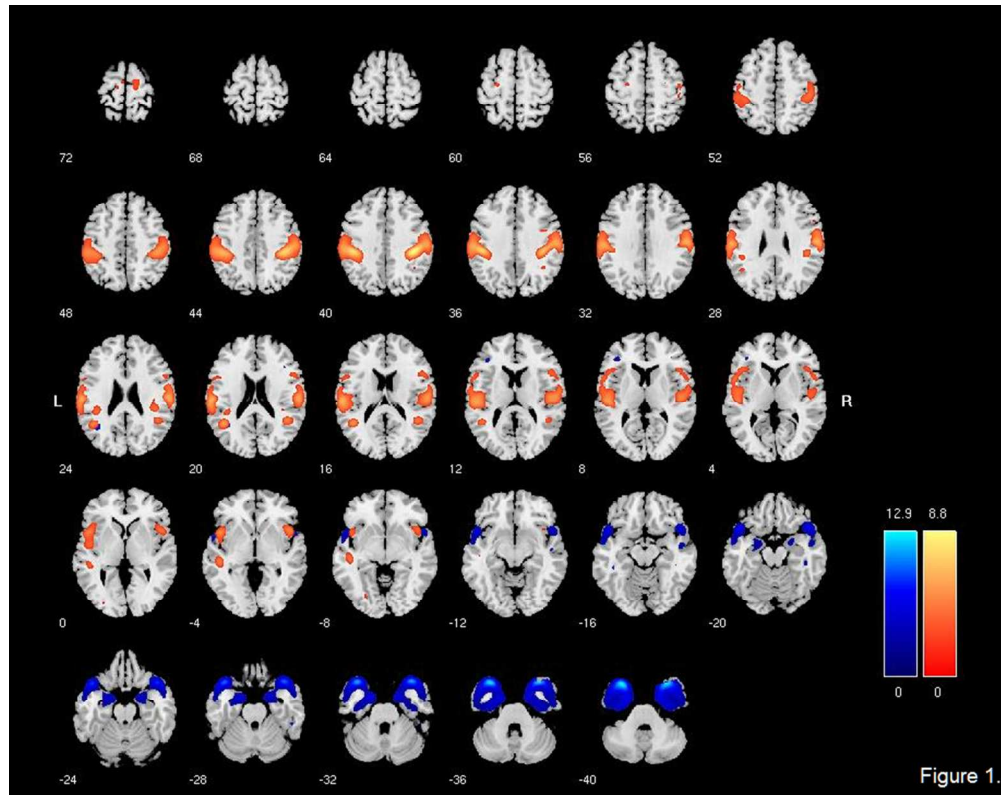


Fig. 1 Series represents a generic human brain, sliced top to bottom in 4mm slices. Colored regions represent the gray matter volume (modulated) associated with a specific component derived from Source Based Morphometry (SBM). Blue represents component areas with higher loading for males ( $t = 14.02$ ,  $p < .001$ ). Orange represents component areas with higher loading for females ( $t = 10.77$ ,  $p < .001$ ). Higher loadings are dimensionally proportional to higher volume in these regions, but do not represent direct differences between males and females (see Figure 3). Colored bars are z-values for the spatial extent of the individual components across subjects.

245x194mm (95 x 95 DPI)

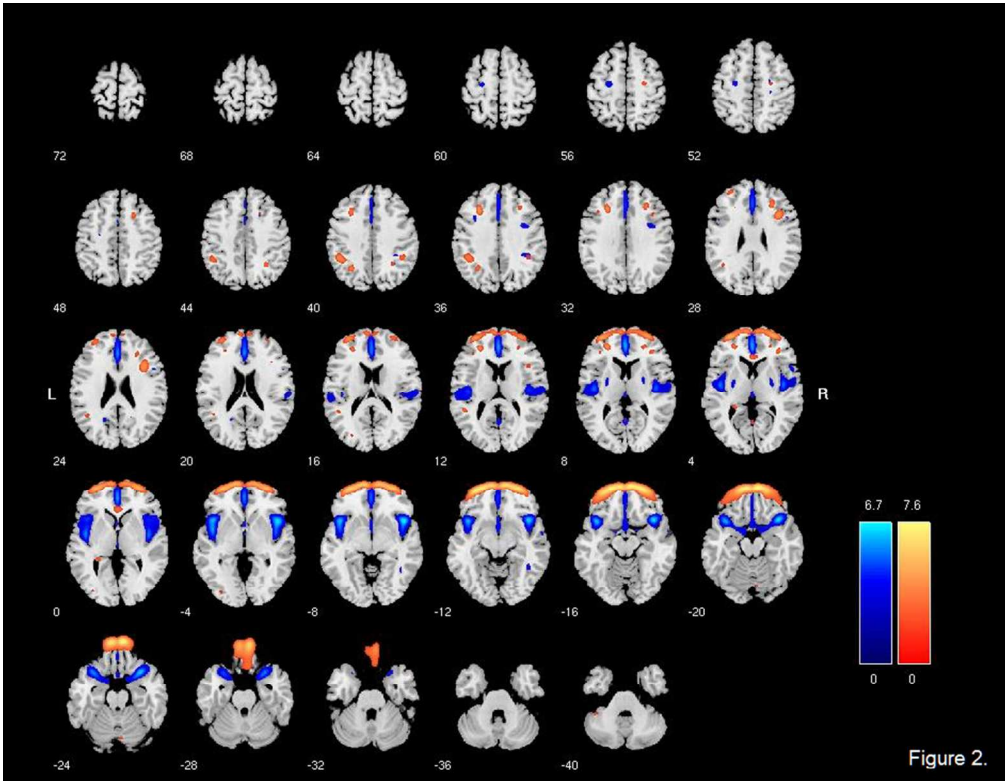


Fig. 2 Series represents a generic human brain, sliced top to bottom in 4mm slices. Colored regions represent the gray matter density (unmodulated) associated with a specific component derived from Source Based Morphometry (SBM). Blue represents component areas with higher loading for males ( $t = 6.02$ ,  $p < .001$ ). Orange represents component areas with higher loading for females ( $t = 4.53$ ,  $p < .001$ ). Higher loadings are dimensionally proportional to higher gray matter concentration in these regions, but do not represent direct differences between males and females (see Figure 3). Colored bars are z-values for the spatial extent of the individual components across subjects.

251x194mm (95 x 95 DPI)



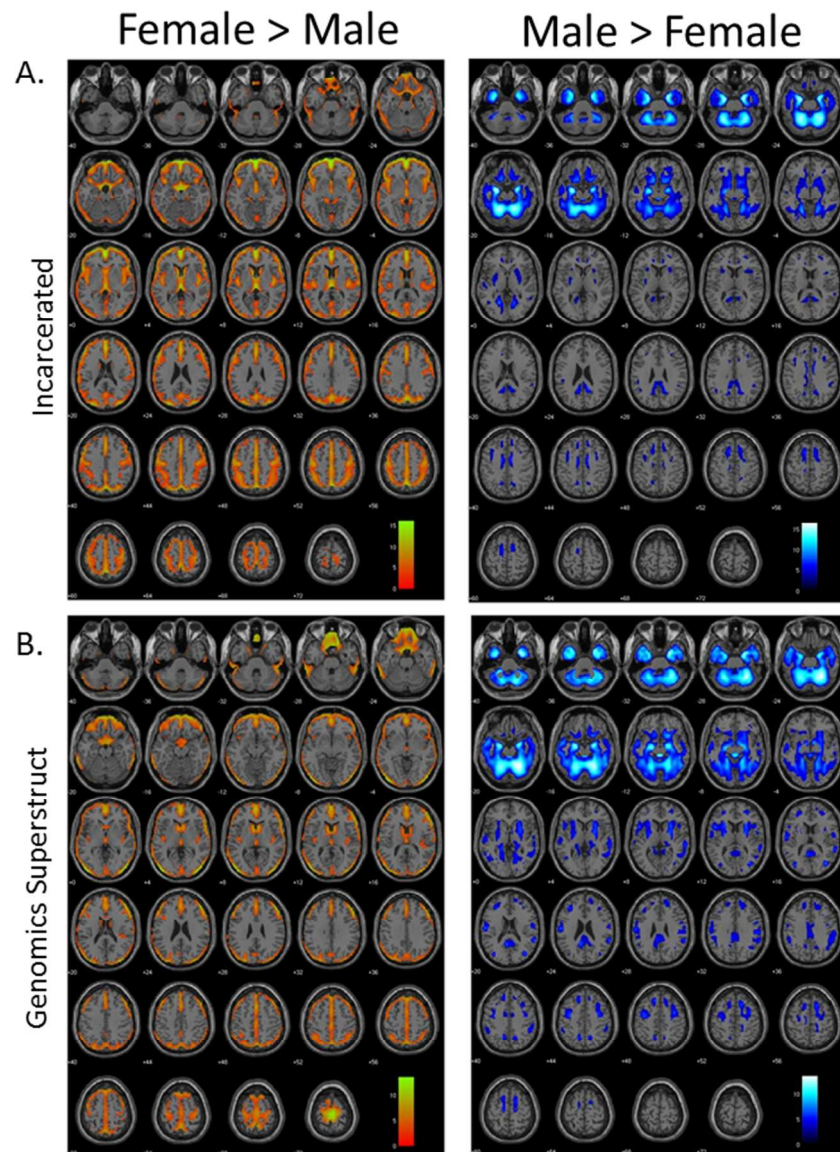


Fig.3 T-tests demonstrating volumetric comparisons between males and females in the incarcerated (A) and non-incarcerated (B: Genomics Superstruct) samples. Orange maps show gray matter relatively larger in females; Blue maps show gray matter relatively larger in males. These figures are intended as a succinct summary of regional gray matter differences for ease of viewing; however, classification models were based on 57 components of gray matter derived from source based morphometry (e.g. Figures 1 & 2).

182x252mm (96 x 96 DPI)