Web-based Supplementary Materials for "Evaluating independent component analyses with an application to resting-state fMRI," by Benjamin B. Risk, David S. Matteson, David Ruppert, Ani Eloyan, and Brian S. Caffo

Web Appendix A: Simulation Studies

A.1 The Infomax algorithm

We are not aware of functions or packages in R that implement the Infomax algorithm (Bell and Sejnowski 1995). We offer an alternative to Matlab code (`http://cnl.salk.edu/~tewon/ICA/code.html`), but with a few modifications that decrease computation time. First, we use the full data (the so-called offline algorithm) in each iteration rather than an online algorithm with batches. Secondly, we use an adaptive method to choose the step size (based upon Bernaards and Jennrich 2005), which speeds up convergence. We also omitted the bias term (intercept) included in the original formulation because we centered our data. R code implementing the Infomax algorithm and example simulations are available in `<EvaluatingICA_Rsources.R>` and `<EvaluatingICA_Examples.R>` in the Supplementary Materials.

A.2 The ProDenICA algorithm

We made small modifications in the simulated data analysis in order to use the R-package ProDenICA. When the IC density was heavy-tailed (e.g., t-distribution with $df = 3$ or

$df = 5$), the algorithm sometimes failed in the density estimation step. These issues were resolved by removing one or more of the most extreme outliers.

It should be noted that the 'restarts' option in the `ProDenICA()` function evaluates the objective function at N random matrices, determines the matrix with the highest negentropy, and then initiates the ProDenICA algorithm with this single matrix. We found that `ProDenICA()` should instead be initiated using multiple random matrices because a single initial value may have a relatively high initial negentropy but be in a basin with a local maximum.

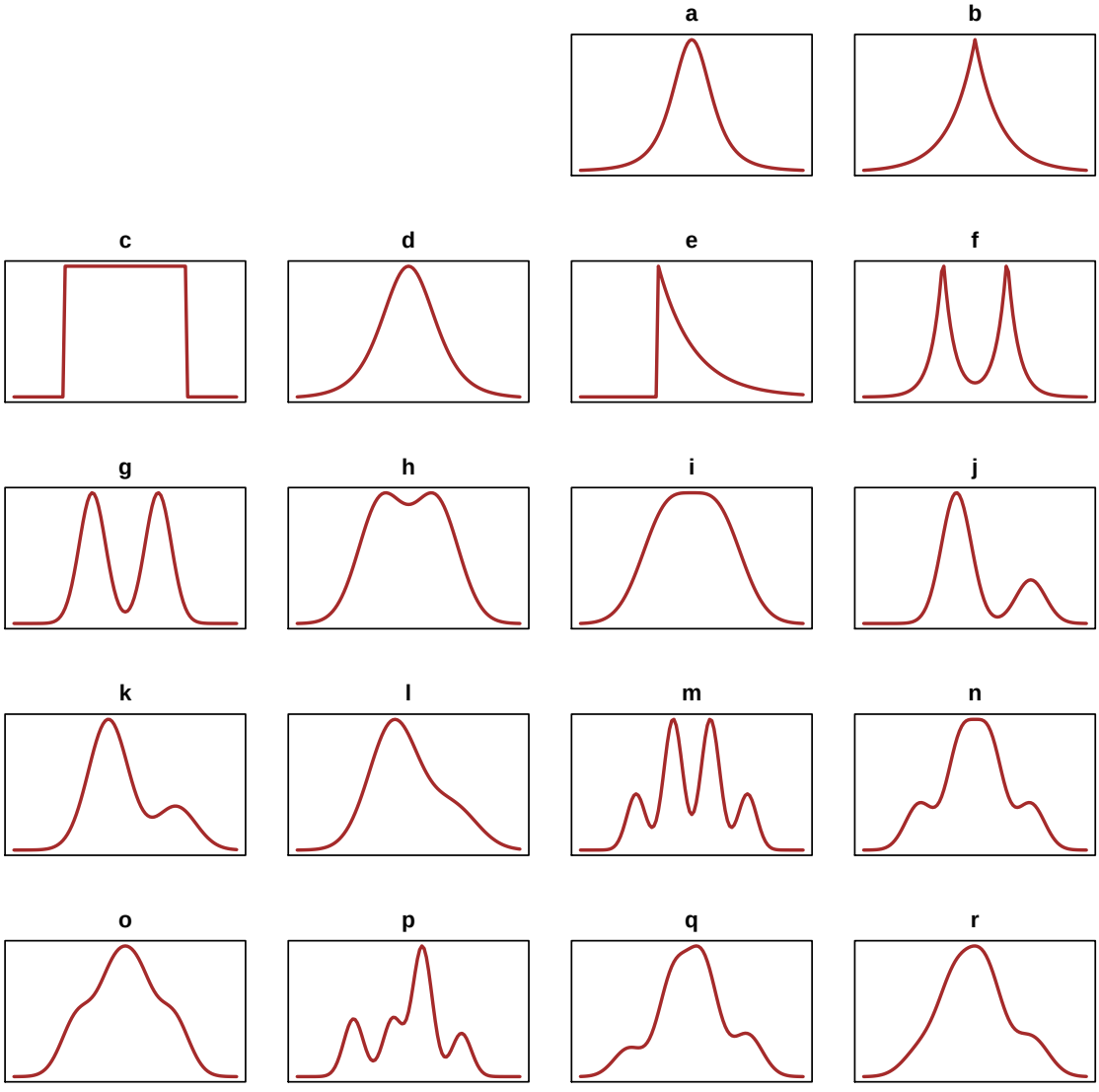
Another issue that arose is that `ProDenICA()` produced an error when using the whitening option with $Q < T_r$. This issue was resolved by supplying `ProDenICA()` with an initial unmixing matrix (rather than relying upon the default).

Lastly, we found that when using the log cosh nonlinearity (`ProDenICA()` provides a function that replicates `fastICA()`), the negentropy measure was not correct; it simply calculated the mean of $\frac{1}{\alpha} \log \cosh(\alpha s)$. It should instead apply the formula in Equation 6 of the manuscript.

A.3 Simulated data

We simulated the mixing matrix $\mathbf{A}$ using the `mixmat()` function from the R package ProDenICA (Hastie and Tibshirani 2010), which ensures the condition number is between 1 and 2 by simulating a $Q \times Q$ matrix with iid entries from a standard normal, taking the SVD, then generating random eigenvalues from the uniform(1,2) distribution, and defining $\mathbf{A}$ as the product of the left eigenvector, these new eigenvalues, and the right eigenvector. We conducted 100 simulations with $V = 1,024$ samples for each component. Twenty-five initial values were used for the iterative methods, where initial values were randomly selected from a latin hypercube using the angular (Givens) parameterization, with $\theta_q \in [0, \pi]$ for $q = 1, \ldots, Q(Q-1)/2 - 1$ and $\theta_{Q(Q-1)/2} \in [0, \pi/2]$. Data were simulated from eighteen distributions using `rjordan()` in the ProDenICA package (Hastie and Tibshirani 2010; Web Figure 1).

Web Figure 1: Distributions used in simulations, which include the t-distribution with df=3, double exponential, uniform, t-distribution with df=5, exponential, a mixture of exponentials, and numerous mixtures of normals. Note that a, b, d, and e are super-Gaussian, while c and f - r are sub-Gaussian.



Web Table 1: The 0.025, 0.500, and 0.975 quantiles of computation times (in seconds) based on 100 simulations with 25 initial values per simulation. Quantiles are based on the pooled sample of 2,500 computation times for all methods except for JADE, which is not initialized with multiple starting values and is consequently based on 100 samples.

Q	Quantile	FastICA	Infomax	JADE	ProDenICA
5	0.025	0.01	1.28	0.02	3.43
5	0.500	0.03	3.19	0.02	5.84
5	0.975	1.58	5.95	0.05	30.67
10	0.025	0.04	5.88	0.10	11.70
10	0.500	0.34	11.69	0.17	28.75
10	0.975	2.85	13.05	0.27	267.23
20	0.025	1.11	18.75	2.41	95.66
20	0.500	7.46	25.36	3.98	544.92
20	0.975	27.07	29.02	10.00	2478.45

A.4 Notes on the minimum distance measure

We adapt the minimum distance (MD) measure (Ilmonen et al. 2010), which was defined for some estimate $\widehat{\mathbf{W}}_{(i)}$ when the true unmixing matrix, $\mathbf{W}$, is known. We apply the measure to two arbitrary square matrices $\mathbf{B}_{(i)}$ and $\mathbf{B}_{(j)}$. Let $\mathcal{P}$ denote the set of $Q \times Q$ signed permutation matrices and $\mathcal{C}$ the set of $Q \times Q$ full-rank diagonal matrices. Then define the set of scaled permutation matrices $\mathcal{K} = \{\mathbf{K} : \mathbf{K} = \mathbf{P}_{\pm}\mathbf{C},\ \forall\ \mathbf{P}_{\pm} \in \mathcal{P},\ \mathbf{C} \in \mathcal{C}\}$. Then the minimum distance measure between two matrices $\mathbf{B}_{(i)}$ and $\mathbf{B}_{(j)}$ is

$$d_{MD}(\mathbf{B}_{(i)}, \mathbf{B}_{(j)}) = \frac{1}{\sqrt{Q-1}} \inf_{\mathbf{K} \in \mathcal{K}} || \mathbf{K}\mathbf{B}_{(i)}\mathbf{B}_{(j)}^{-1} - \mathbf{I}_d ||_F$$

where $|| \cdot ||_F$ denotes the Frobenius norm. Code implementing this measure is available in the R package *JADE* (Nordhausen et al. 2011).

A.5 Computation times

We conducted our simulations on a cluster of 28 Dell PowerEdge 2650 servers with 8 processors per server, where each processor was 2.66 GHz. We used the R package *snow* (Tierney et al. 2011) to conduct simulations in parallel. Computation times are presented in Web Table 1.

Web Appendix B: Matching ICs

Our approach to matching ICs follows a modification of the Hungarian (Kuhn-Munkres) algorithm (Tichavsky and Koldovsky 2004), and here we describe the modification in detail. Suppose we want to compare $\widehat{\mathbf{S}}^k_{(i)} \in \mathbb{R}^{V\times Q}$ and $\widehat{\mathbf{S}}^l_{(j)} \in \mathbb{R}^{V\times Q}$, the ith estimate from method k and the jth estimate from method l. Hereafter, we drop the k and l superscripts to simplify notation, noting that the estimates may or may not be from the same method. Assume that $\widehat{\mathbf{S}}_{(i)}$ is in canonical form, as defined in Section 4.2. We refer to the canonically ordered $\widehat{\mathbf{S}}_{(i)}$ as the template. Let $\widehat{\mathbf{S}}_{(i),q}$ be the qth column of $\widehat{\mathbf{S}}_{(i)}$ and $\widehat{\mathbf{S}}_{(j),r}$ be the rth column from $\widehat{\mathbf{S}}_{(j)}$, and let $||\cdot||$ denote the Euclidean norm. Create a $Q \times Q$ distance (cost) matrix $\mathbf{C}$ between the components with elements

$$c_{q,r} = \min\left(||\widehat{\mathbf{S}}_{(i),q} - \widehat{\mathbf{S}}_{(j),r}||, ||\widehat{\mathbf{S}}_{(i),q} + \widehat{\mathbf{S}}_{(j),r}||\right),$$

and define the matrix $\mathbf{B}$ with

$$b_{q,r} = \begin{cases} -1 & \text{if } \min\left(||\widehat{\mathbf{S}}_{(i),q} - \widehat{\mathbf{S}}_{(j),r}||, ||\widehat{\mathbf{S}}_{(i),q} + \widehat{\mathbf{S}}_{(j),r}||\right) = ||\widehat{\mathbf{S}}_{(i),q} + \widehat{\mathbf{S}}_{(j),r}||, \\ 1 & \text{if } \min\left(||\widehat{\mathbf{S}}_{(i),q} - \widehat{\mathbf{S}}_{(j),r}||, ||\widehat{\mathbf{S}}_{(i),q} + \widehat{\mathbf{S}}_{(j),r}||\right) = ||\widehat{\mathbf{S}}_{(i),q} - \widehat{\mathbf{S}}_{(j),r}||. \end{cases}$$

Let $\mathbf{S}$ be the set of all permutations of the integers 1 to Q, where for some $\sigma \in \mathbf{S}$, we denote the permutation $\sigma = (\sigma(1), \ldots, \sigma(Q))$. We then use the Hungarian algorithm (Kuhn 1955) to identify the set such that

$$\sigma^* = \underset{\sigma\in\mathcal{S}}{\operatorname{argmin}} \sum_{q=1}^{Q} c_{q,\sigma(q)}.$$

Then define the signed permutation matrix $\mathbf{P}_1$ with entries $p_{q,a_q} = b_{q,a_q}$ at row q and column a_q, and 0 otherwise. Note that $\mathbf{P}_1$ is equivalent to $\underset{\mathbf{P}_\pm\in\mathcal{P}}{\operatorname{argmin}} ||\ \widehat{\mathbf{S}}_{(i)} - \widehat{\mathbf{S}}_{(j)}\mathbf{P}'_\pm\ ||_F$.

The method used here to match ICs creates a one-to-one mapping of components. Note that when multiple ICs are being compared, the matching algorithm may be sensitive to the choice of template. In our application, we found that using the estimates from JADE, Infomax, or ProDenICA as the template with one-at-a-time matching resulted in the same ordering as using the FastICA estimate as the template. In situations in which ICs from more than two estimates differ greatly, a method to simultaneously match all ICs could be

Web Table 2. Subject diagnosis by site in the ADHD-200 Sample: Typ=Typically Developing; ADHD-C=ADHD-Combined; ADHD-H/Im=ADHD-Hyperactive and Impulsive; ADHD-In=ADHD-Inattentive; WH= Withheld.

Site	Typ	ADHD-C	ADHD-H/Im	ADHD-In	WH
Bradley Hospital/Brown University	0	0	0	0	26
Kennedy Krieger Institute	61	16	1	5	11
NeuroIMAGE Sample	23	18	6	1	25
NYU Child Study Center	99	77	2	44	41
Oregon Health & Science University	42	23	2	12	34
Peking University	116	29	0	49	51
University of Pittsburgh	89	0	0	0	9
Washington University in St. Louis	61	0	0	0	0
Total	491	163	11	111	197

pursued.

Web Appendix C: Group ICA of the ADHD-200 Sample

C.1 Resting-state fMRI dataset

Data were selected for analysis from the ADHD-200 Data Sample, which consists of rs-fMRI data from children and adolescents (ages 7-21) from 8 independent sites comprising 491 typically developing subjects and 285 that were diagnosed with ADHD (Web Table 1). Subjects were diagnosed with three ADHD subtypes: Inattentive; Hyperactive and Impulsive; and Combined (Hyperactive/Impulsive and Inattentive). However, there were only a total of 11 subjects with ADHD-Inattentive, and half the sites did not have subjects with this diagnosis.

We restricted our analysis to (1) subjects with no recorded history of drug therapy; (2) subjects that were right-hand dominant; (3) images with no quality control flags; and (4) subjects that were either ADHD-Combined or ADHD-Inattentive (but not ADHD-Hyperactive and Impulsive). Subjects were classified using either (1) the ADHD Rating Scale IV, (2) Conner's Parent Rating Scale-Revised (Long Version), or (3) Conner's Rating Scale, 3rd edition. Within these scales, there was a small degree of overlap in the intermediate values between subjects diagnosed as typically developing and subjects diagnosed with ADHD, whereas individuals with low values were strictly labeled typically developing

Web Table 3. Subjects used in analysis. Typ=Typically Developing; ADHD-C=ADHD-Combined; ADHD-In=ADHD-Inattentive.

Site	Typ	ADHD-C	ADHD-H/Im	ADHD-In	WH
Peking University	86	13	0	19	0
Kennedy Krieger Institute	40	7	0	3	0
NYU Child Study Center	56	16	0	11	0
Oregon Health & Science University	24	8	0	1	0
Total	206	44	0	34	0

and individuals with high values were strictly diagnosed with ADHD. We excluded subjects with scores that we deemed borderline, that is, both control and ADHD subjects that were near the threshold at which ADHD was diagnosed. Specifically, we excluded subjects with ADHD Rating Scale IV values between 36 and 45; Conner's Parent Rating Scale-Revised (Long Version) between 56 and 65; or Conner's Rating Scale between 55 and 66 (Web Table 2).

Details of the primary image processing pipeline were previously reported (Section 2.1, Eloyan et al. 2012). Processing followed the functional connectome processing scripts on the FCP/INDI site (Mennes et al. 2012). In addition, we aggregated the MNI 152 T1 3 mm template to result in $6 \times 6 \times 6$ mm voxels. We retained the $6 \times 6 \times 6$ mm voxels for which all eight of the voxels in the MNI 152 T1 3mm template were brain tissue. This resulted in $V = 7,825$ for all subsequent analyses. For subjects in which there were multiple scanning sessions, we only used the first session.

We also used our own whitening function to produce the input data for all algorithms, as provided in `<EvaluatingICA_Rsources.R>`. This ensured that $\hat{\mathbf{W}}$ and $\hat{\mathbf{S}}$ were always defined equivalently. Note that the functions `fastICA()` and `JADE()` automatically whiten data; consequently, we modified the source code to prevent additional whitening.

C.2 Differences Between Algorithms

We compared the unmixing matrices from FastICA, Infomax, JADE, and ProDenICA using the MD measure, the Amari measure, and the Frobenius distance between matched unmixing matrices. To aid in our interpretation of the magnitude of differences between mixing matrices, we simulated the distribution of these three measures for randomly generated

orthogonal matrices using two methods. First, orthogonal matrices were generated with columns equal to the eigenvectors from the spectral decompositions of randomly generated matrices following a Wishart distribution with covariance equal to the identity matrix and V degrees of freedom. Second, we simulated uniformly distributed Givens rotation angles $\theta_i \in [-\pi, \pi]$ for $i = 1, \ldots, Q(Q-1)/2$, and then converted the angles to orthogonal matrices (Web Table 4).

Web Table 4. Distance and measures between unmixing matrices by method for the rs-fMRI study. Here, the SVD mixing matrix is taken to be the identity matrix. MD = Minimum Distance measure. Mean and 1% Wishart denote the mean and 1% quantiles, respectively, of each measure from matrices randomly generated via the SVD of iid Wishart matrices. Mean and 1% unif denote the corresponding statistics for matrices generated from the angular parametrization of orthogonal matrices with angles uniformly distributed in $[-\pi, \pi]$.

Method.1	Method.2	Amari	MD	Frobenius
Mean: Wishart 1	Wishart 2	0.35	0.90	6.31
1%: Wishart 1	Wishart 2	0.31	0.88	5.92
Mean: Unif 1	Unif 2	0.26	0.85	6.32
1%: Unif 1	Unif 2	0.22	0.80	5.76
SVD	fastICA	0.36	0.91	6.30
SVD	Infomax	0.36	0.91	6.33
SVD	JADE	0.35	0.90	6.30
SVD	ProDenICA	0.33	0.89	6.29
FastICA	Infomax	0.01	0.07	0.29
FastICA	JADE	0.06	0.38	1.75
FastICA	ProDenICA	0.06	0.41	1.89
Infomax	JADE	0.06	0.39	1.80
Infomax	ProDenICA	0.06	0.42	1.93
JADE	ProDenICA	0.07	0.41	1.85

In Web Table 5, we present false discovery rate (FDR) adjusted p-values from two-sample Kolmogorov-Smirnov tests for equality in distribution between ICs estimated using the SVD, FastICA, Infomax, and ProDenICA. In multiple hypothesis testing, the FDR is the expected proportion of false positives among the rejected null hypotheses, and controlling the FDR leads to more powerful testing procedures than controlling the family-wise error rate (Benjamini and Hochberg 1995). For each p-value, we calculated an FDR-adjusted p-value, called a *q-value* (Storey 2002): let G denote the total number of tests and $p_{(g)}$

denote the gth order statistic from the set of all G p-values, and define the q-value

$$p^*_{(g)} = \min\left(\frac{G}{g}p_{(g)}, p^*_{(g+1)}, \ldots, p^*_{(G)}, 1\right).$$

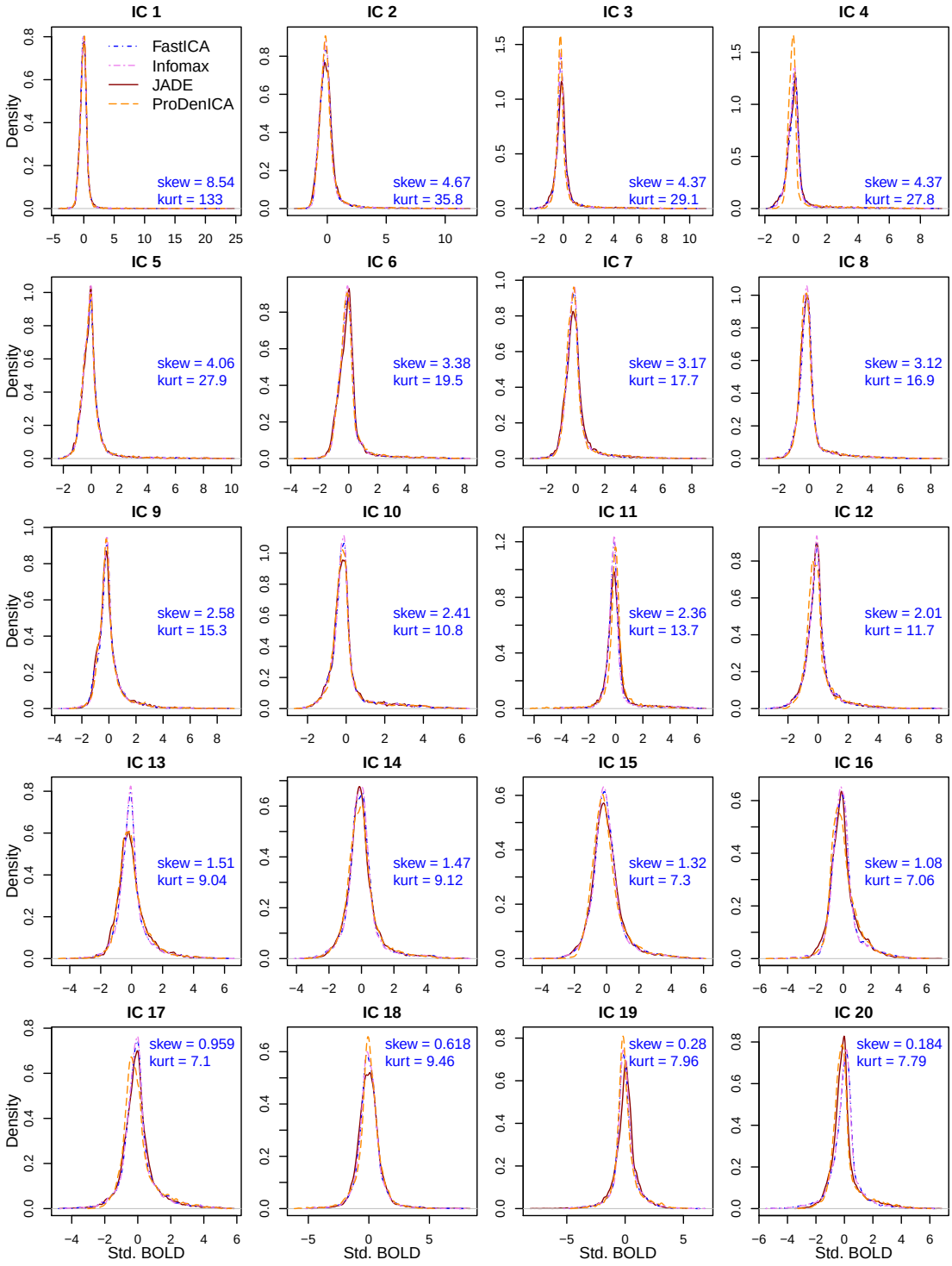
In typical applications, $p^*_{(g)}$ is an estimate of the minimum proportion of false positives given that at least one rejection occurs, where the minimum is taken over all rejection regions containing $[0, p_{(g)}]$. Here, we use the FDR-adjusted p-values as a measure of the difference between IC distributions since the test statistics were based on spatially dependent data.

Web Table 5. FDR-adjusted p-values from two-sample Kolmogorov-Smirnov statistics. Blank entries indicate FDR-adjusted $p < 0.0001$.

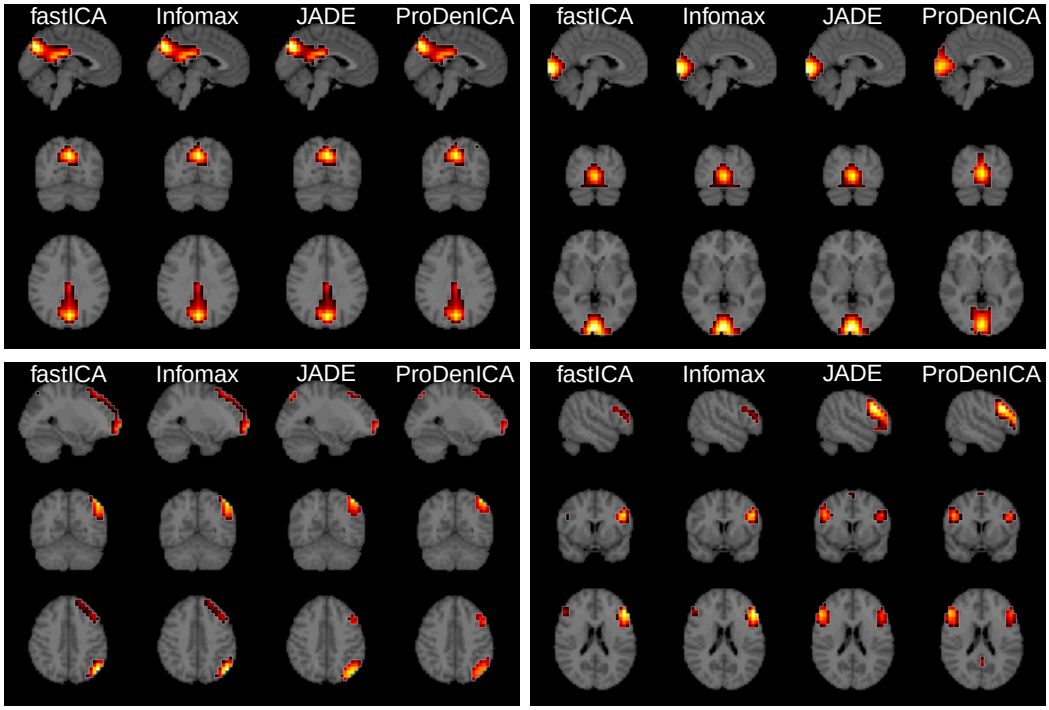
Method1	Method2	IC 1	IC 2	IC 3	IC 4	IC 5	IC 6	IC 7	IC 8	IC 9	IC 10
SVD	FastICA										
SVD	Infomax										
SVD	JADE										
SVD	ProDenICA										
FastICA	Infomax	0.9826	0.2733	0.1277		0.0556	0.5650	0.3543	0.4036	0.1105	0.9481
FastICA	JADE	0.6165	0.0101		0.4788	0.2421	0.0001	0.0004	0.0222	0.0003	0.0129
FastICA	ProDenICA	0.0658	0.0166			0.0451	0.1277	0.0002		0.0053	0.0129
Infomax	JADE	0.4688				0.1574			0.0556	0.0004	0.0024
Infomax	ProDenICA	0.1370	0.2660			0.2354	0.1811	0.0005		0.0254	0.0027
JADE	ProDenICA	0.2807				0.0254			0.0002	0.0265	0.1415
Method1	Method2	IC 11	IC 12	IC 13	IC 14	IC 15	IC 16	IC 17	IC 18	IC 19	IC 20
SVD	FastICA								0.0004		
SVD	Infomax								0.0003		
SVD	JADE								0.0002		
SVD	ProDenICA							0.0018			
FastICA	Infomax	0.0878	0.4890	0.3943	0.3851	0.9826	0.4225	0.7906	0.9826	0.2867	0.4130
FastICA	JADE		0.2136		0.0380	0.1068	0.0101	0.1866	0.0006		
FastICA	ProDenICA										
Infomax	JADE		0.9826		0.0112	0.0433	0.0002	0.0348	0.0002		
Infomax	ProDenICA										
JADE	ProDenICA			0.2867	0.0304						

We also present density plots for each IC and each method. Densities were estimated using a Gaussian kernel. For each component, a bandwidth was determined for the estimate from FastICA, Infomax, JADE, and ProDenICA, respectively, using the method of Sheather and Jones (1991), and then these four bandwidths were averaged, and finally the densities were estimated with bandwidth fixed at this average. Thus, for a given component, the densities for each of the methods were estimated using the same bandwidth.

Web Figure 2. Density plots of ICs for FastICA, Infomax, JADE, and ProDenICA. Values on the x-axis correspond to the standardized BOLD signal. The sample skewness and kurtosis from the FastICA estimates are included in the plot area.



Web Figure 3. Estimated ICs. Clockwise from the top-left: IC 3 (parts of default network), IC 4 (parts of the visual cortex), IC 13 (strong lateralization for FastICA and Infomax but not JADE and ProDenICA), and IC 20 (strong lateralization in all methods; areas associated with memory).

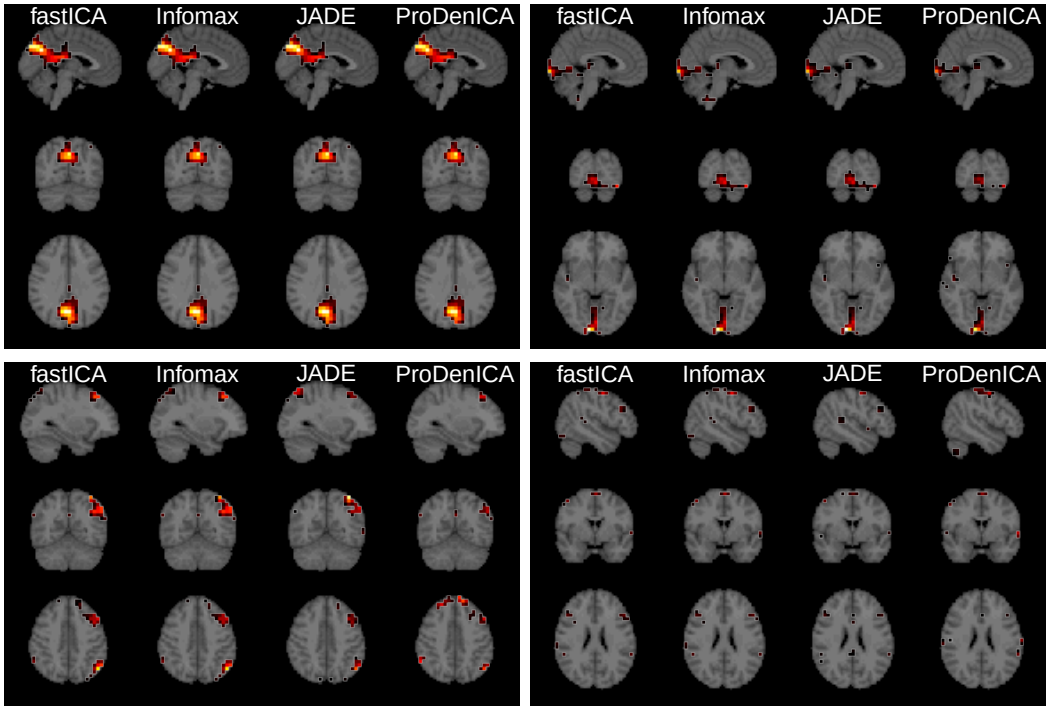


Selected resting-state networks

Web Figure 3 presents images for selected ICs from the group ICA of the ADHD-200 Data Sample. Images were thresholded to retain voxels with values greater than the 97.5% quantile. Slices were chosen to approximately maximize the number of visible activated voxels.

We estimated ICs from a single individual randomly chosen from the ADHD-200 data (subject ID 3446674.1.1.pek2). We matched the FastICA estimates from this individual to the skewness-ordered FastICA estimates of the group ICs, and then matched the ICs from Infomax, JADE, and ProDenICA to these re-ordered FastICA results. Selected ICs are presented in Web Figure 4.

Web Figure 4. Estimated ICs for a single subject randomly chosen from the ADHD200 dataset (subject ID 3446674.1.1.pek2). Clockwise from the top-left: IC 3 (parts of default network), IC 4 (medial areas of the visual cortex), IC 13, and IC 20 (strong lateralization in all methods; areas associated with memory).



References

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation* **7,** 1129–1159.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* pages 289–300.

Bernaards, C. A. and Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement* **65,** 676–696.

Eloyan, A., Muschelli, J., Nebel, M. B., Liu, H., Han, F., Zhao, T., Barber, A., Joel, S., Pekar, J. J., Mostofsky, S., and Caffo, B. S. (2012). Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. *Frontiers in Systems Neuroscience* **6,** 1–9.

Hastie, T. and Tibshirani, R. (2010). *ProDenICA: Product Density Estimation for ICA using tilted Gaussian density estimates*. R package version 1.0.

Ilmonen, P., Nordhausen, K., Oja, H., and Ollila, E. (2010). A new performance index for ICA: properties, computation and asymptotic analysis. *Latent Variable Analysis and Signal Separation* pages 229–236.

Kuhn, H. W. (1955). The Hungarian Method for the assignment problem. *Naval Research Logistics Quarterly* **2,** 83 – 97.

Mennes, M., Biswal, B., Castellanos, F. X., and Milham, M. P. (2012). Making data sharing work: The FCP/INDI experience. *NeuroImage* .

Nordhausen, K., Cardoso, J. F., Oja, H., and Ollila, E. (2011). *JADE: JADE and ICA performance criteria*. R package version 1.0-4.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 683–690.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 479–498.

Tichavsky, P. and Koldovsky, Z. (2004). Optimal pairing of signal components separated by blind techniques. *Signal Processing Letters, IEEE* **11,** 119–122.

Tierney, L., Rossini, A. J., Li, N., and Sevcikova, H. (2011). *snow: Simple Network of Workstations*. R package version 0.3-8.