NeuroImage: Clinical 9 (2015) 244–252

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

Functional neuroimaging of visuospatial working memory tasks enables accurate detection of attention deficit and hyperactivity disorder

Rubi Hammer[a,b,*], Gillian E. Cooke[a,c], Mark A. Stein[d], James R. Booth[a,b,e]

[a]Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA
[b]Interdepartmental Neuroscience Program, Northwestern University, Evanston, IL, USA
[c]Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana-Champaign, IL, USA
[d]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, USA
[e]Department of Communication Sciences and Disorders, The University of Texas at Austin, Austin, TX, USA

ARTICLE INFO

Article history:
Received 29 July 2015
Received in revised form 21 August 2015
Accepted 25 August 2015
Available online 1 September 2015

Keywords:
ADHD diagnosis
Neurobiological marker
Visuospatial working memory
Reward processing
Feedback processing

ABSTRACT

Finding neurobiological markers for neurodevelopmental disorders, such as attention deficit and hyperactivity disorder (ADHD), is a major objective of clinicians and neuroscientists. We examined if functional Magnetic Resonance Imaging (fMRI) data from a few distinct visuospatial working memory (VSWM) tasks enables accurately detecting cases with ADHD. We tested 20 boys with ADHD combined type and 20 typically developed (TD) boys in four VSWM tasks that differed in feedback availability (feedback, no-feedback) and reward size (large, small). We used a multimodal analysis based on brain activity in 16 regions of interest, significantly activated or deactivated in the four VSWM tasks (based on the entire participants' sample). Dimensionality of the data was reduced into 10 principal components that were used as the input variables to a logistic regression classifier. fMRI data from the four VSWM tasks enabled a classification accuracy of 92.5%, with high predicted ADHD probability values for most clinical cases, and low predicted ADHD probabilities for most TDs. This accuracy level was higher than those achieved by using the fMRI data of any single task, or the respective behavioral data. This indicates that task-based fMRI data acquired while participants perform a few distinct VSWM tasks enables improved detection of clinical cases.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Attention-deficit and hyperactivity-disorder (ADHD) affects 5–8% of the worldwide childhood population, it has high comorbidity and it shares symptoms with other behavioral and emotional disorders (Boyle et al., 2011; Larson et al., 2011). ADHD prevalence rates vary significantly between and within countries, and depend on the ascertainment method and criteria utilized. Of particular concern is the marked variability in diagnostic rates of ADHD in developed countries (Getahun et al., 2013). This may reflect increased awareness of teachers and parents to symptoms of ADHD in communities with a focus on education and adequate health care. However, even within communities there is marked variability in rates of ADHD due to the subjective nature of primarily descriptive (symptom-based) diagnostic procedures (Asherson et al., 2012; Polanczyk et al., 2014; see also recent reports by the US Center for Disease Control and Prevention). Such diagnostic heterogeneity and its poor reliability have hampered efforts to determine the pathophysiology of ADHD (Morgan et al., 2013).

Attempts have been made to find neuroimaging-based biomarkers of ADHD using structural Magnetic Resonance Imaging (Johnston et al., 2014; Lim et al., 2013; Peng et al., 2013), resting-state functional MRI (Hoekzema et al., 2014; Tomasi and Volkow, 2014), fMRI data acquired during a single cognitive task (Hale et al., 2015; Hart et al., 2014a; Hart et al., 2014b), or combinations of some of these techniques (Anderson et al., 2014; ADHD-200 Consortium, 2012; Iannaccone et al., 2015). These studies provide mixed insights regarding the possible underlying neurocognitive deficit of ADHD, and they report moderate classification accuracies (rarely exceeding 80%), insufficient for clinical diagnosis.

ADHD is characterized by behavioral symptoms and multiple cognitive deficits associated with context dependent abnormal patterns of neural activity distributed across multiple brain regions. Consequently, ADHD is unlikely to be characterized by localized structural brain abnormalities robust enough to be detected using available structural imaging techniques. It is also unlikely that significant functional brain abnormalities characterizing ADHD would be evident regardless of the mental state of the test subject, during the scan (Castellanos and Proal, 2012; Hammer et al., 2015).

Key characteristics of ADHD include poor working memory, greater reliance on external feedback, and abnormal reward processing. These

* Corresponding author at: Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208-2952, USA. Tel.: +1 650 799 3790.

http://dx.doi.org/10.1016/j.nicl.2015.08.015
2213-1582/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

are associated with altered patterns of activity in distinct brain networks: (i) attention and working memory, comprising the dorsolateral prefrontal cortex, parietal cortex and temporal cortices (Burgess et al., 2010; Ehlis et al., 2008; Vance et al., 2007); (ii) executive functions and cognitive control (including feedback processing and response selection), comprising the dorsal, medial and ventral frontal cortices (Booth et al., 2005; Clark et al., 2007; Sonuga-Barke and Fairchild, 2012); and (iii) reward-processing, comprising the orbitofrontal cortex, anterior cingulate cortex and basal ganglia (del Campo et al., 2013; Plichta et al., 2009; Volkow et al., 2009). Notably, there are substantial individual differences in reactions to reward and feedback manipulation in ADHD (Demurie et al., 2011; Dovis et al., 2015; Hammer et al., 2015; Plichta and Scheres, 2014; van der Schaaf et al., 2013).

Accordingly, we hypothesized that using fMRI data from an ensemble of visuospatial working memory (VSWM) tasks that differ in the motivational context (determined by the availability of reward and feedback) would increase the odds that abnormal patterns of brain activity would be reliably detected in a larger proportion of ADHD cases. We expected this to enable more accurate detection of ADHD cases than the accuracies obtained by using fMRI data acquired from any single VSWM task, or the accuracy based on the participant's respective behavioral performance.

Given the central role of VSWM in ADHD, we compared fMRI data from four distinct VSWM tasks in boys with ADHD and typically developed (TD) boys. Tasks differed in the availability of trial-by-trial feedback (feedback versus no-feedback) and the participant's expectation for significant monetary reward (large versus small). All tasks required tracking the spatial location of letters while ignoring the letters' identity, and executing timely responses. The manipulation of feedback and reward provided different motivational contexts, each requiring somewhat distinct executive skills found to be impaired in children with ADHD.

We used a multimodal analysis based on relatively few brain regions of interest, discovered using an independent univariate analysis (Hammer et al., 2015; Morris et al., 2012; Mulligan et al., 2011; Thoma and Henson, 2011). Using a sparse principal component analysis, we further reduced the number of variables that were provided as input to the classifier. This substantially limited the odds of discovering an overfitted ADHD classification model, and simplified the interpretation of the discovered model. We used a logistic regression (LR) classifier, which directly models the class conditional probabilities for each case (i.e., calculating the predicted probability that a given child has ADHD) by attempting to find a model allowing a decisive classification of as many cases as possible. This attribute is important in clinical settings where we aim to find a model that allows classification with confidence (Nouretdinov et al., 2011).

2. Materials and methods

2.1. Participants

Twenty boys with a diagnosis of ADHD combined-type (mean age in years = 10.42 ± SD = 0.80) and 20 typically developed (TD) boys (10.96 ± 0.91) participated in the experiment. Participants gave their informed consent (and parental consent) in accordance with the policies of the Institutional Review Board (IRB) at Northwestern University. At the time of the fMRI scanning session, ADHD youth were weaned off stimulant medication for at least 24 h (12 participants used prescribed stimulants on a regular basis during the time they participated in this study). ADHD diagnoses were based on exceeding the clinical cut offs on the ADHD rating scale (DuPaul et al., 1998) and the semi-structured diagnostic interview, the Kiddie Schedule for Affective Disorders and Schizophrenia for School Aged Children: Present and Lifetime (K-SADS-PL) version (Kaufman et al., 1997). A psychologist specializing in ADHD reviewed all diagnoses. Mean total ADHD score was higher in the ADHD group than in the TD group (Supplemental Table 1).

Participants were excluded if they had been diagnosed with a neurological disorder or were treated medically for a comorbid psychiatric disorder. All participants were right-handed native English speakers with normal or corrected to normal vision. Mean full-scale IQ scores were within normal range in both groups, however the average IQ in the ADHD group was lower than the average IQ in the TD group (Supplemental Table 1), which is not uncommon in ADHD studies (e.g., Hart et al., 2014a).

2.2. Visuospatial working memory 2-back tasks

Participants performed four VSWM 2-back tasks while being scanned in an fMRI scanner. Each task was 48 trials long. Before and after each 2-back task the participant performed a fixation task where he was asked to press a key whenever the fixation-cross changed its color (which happened in 4/12 of the trials). Brain activity from the fixation tasks was used as baseline. E-Prime ® 2.0 (Psychology Software Tools, INC.) was used for stimuli presentation and for recording participants' responses (for more details about the experimental design, see Hammer et al., 2015).

The two independent factors in the study were reward size (large-reward versus small-reward) and presence of trial-by-trial feedback (no-feedback versus feedback). In an earlier practice session, and at the beginning of the scanning session, each participant was instructed that the reward for each correct decision in the large-reward task was 10 times larger than in the small-reward task. In the trial-by-trial feedback task, each key-press was followed by either a green square (indicating a correct decision) or a red square (indicating an incorrect decision) presented in the center of the screen. In the no-feedback task, the participant was informed about his overall performance only after concluding the task (Fig. 1). Order of the four tasks was counterbalanced across participants.

2.3. MRI and fMRI data acquisition

Imaging data were acquired on a 3.0 Tesla Siemens Tim Trio scanner using a 12-channel head coil. Gradient echo localizer images were acquired to determine the placement of the functional slices. A susceptibility weighted single-shot EPI (echo planar imaging) method with BOLD (blood oxygenation level-dependent) was used for functional image acquisition with the following scan parameters: TR = 2000 ms, TE = 20 ms, flip angle = 80°, matrix size = 128 × 120, field of view = 220 × 206.3 mm, slice thickness = 3 mm (0.48 mm gap), and number of slices = 32 (an effective functional voxel size of 2 × 2 × 4 mm). A total of 145 images (TRs) were recorded for each scan. Slices were acquired in an interleaved manner. A high resolution, T1 weighted 3D image was also acquired with the following parameters: TR = 2300 ms, TE = 3.36 ms, flip angle = 9° matrix size = 256 × 256, field of view = 256 mm, slice thickness = 1 mm, and number of slices = 160. The acquisition of the anatomical scan took approximately 9 min. Prior to the scanning session children were trained in a mock scanner. This enabled confirming that the participant is capable of keeping his head still for the duration of the scanning session. To minimize head movements in the scanner, gaps between the participant's head and the head-coil were filled with memory foam.

2.4. Image preprocessing

Data analysis was performed using MathWorks® Matlab, SPM8 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, London, UK), and IBM® SPSS. Preprocessing involved: (i) slice timing; (ii) realignment of all functional images to the 24th image. (iii) Co-registration of the functional and anatomical images; (iv) Normalization of the T1 image to the MNI305 template image, which is

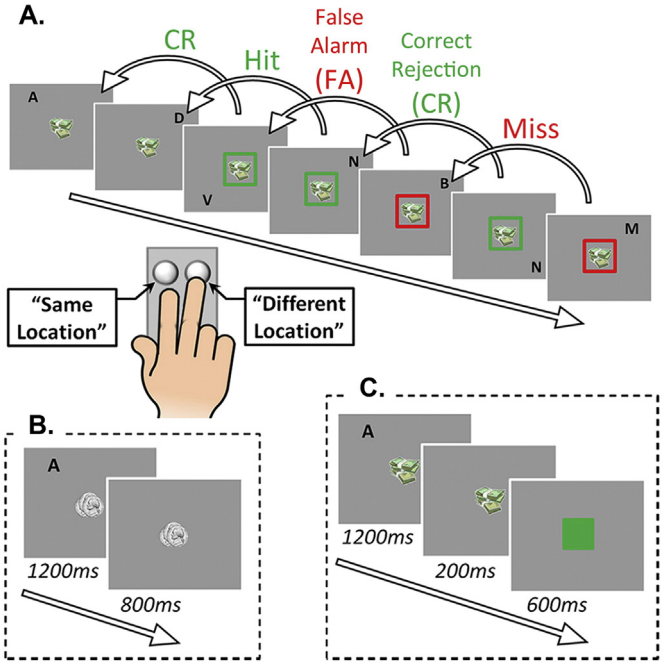

Fig. 1. (A) An illustration of several trials in a 2-back task with large-reward and trial-by-trial feedback. The participants made their decision using the two keys of a response box. The four possible responses were: (i) hit — correctly responding that the location of the current letter was identical to the 2-back letter; (ii) correct rejection (CR) — correctly responding that the location of the current letter was different from the 2-back letter; (iii) false alarm (FA) — incorrectly responding that the location of the current letter was identical to the 2-back letter; and (iv) miss — incorrectly responding that the location of the current letter was different from the 2-back letter when it was actually the same. (B) A single trial in a small-reward (symbolized to the participants by a picture of coins) no-feedback task. The target letter was presented for 1200 ms. Letter presentation was followed by 800 ms presentation of the Reward-Size symbol only. (C) A single trial in a large-reward (symbolized by a picture of dollar bills) trial-by-trial feedback task. Letter presentation was followed by 200 ms presentation of the Reward-Size symbol, which was followed by the presentation of feedback for 600 ms (e.g., a green square indicating a correct response).

most commonly used also for analyzing fMRI data of pediatric populations (e.g., Burgund et al., 2002; Ghosh et al., 2010; Peters et al., 2014; Zhang et al., 2015). Linear and non-linear normalization parameters were then applied to the functional images. (v) 4 × 4 × 8 mm full width half maximum (FWHM) Gaussian kernel smoothing. (vi) We confirmed that movement was kept below 4 mm (in any of the x, y, or z dimensions) within a scan using the ArtRepair software. Images (up to 9 per scan) were realigned in ArtRepair, using interpolated values from the two adjacent non-outlier images. Participants with extensive head movements were rescanned (or excluded). For subsequent general linear model (GLM) analyses, the excluded noisy images were deweighted. As reported in Supplemental Table 2, the two groups did not differ in patterns of head movements, and the replacement of outlier images primarily enabled reducing the signal to noise ratio in the fMRI data of both groups. In order to further reduce within scan variability in neural activity, only trials in which the participant responded correctly (hit and correct rejection) were modeled, with onset time-locked to the beginning of each trial (Calhoun et al., 2005; Demir et al., 2014; Hammer et al., 2015; Puschmann et al., 2013). This had only a small quantitative impact on the reported findings. (vii) A high pass filter with a cut-off of 256 s was applied.

2.5. MRI quality control and head movements estimate

Structural and functional brain images were inspected and found not to have significant image artifacts. We confirmed that for each participant maximal head displacement (the distance between the two most distant fMRI images within a scan) in all translational axes was no larger than the size of a voxel (i.e., 4 mm). There were no significant between-groups differences in any head translation or rotation axes, all $p > 0.25$ (Supplemental Table 2). We also confirmed that it is unlikely that head movements underlie the ADHD classification results we got based on the brain activity data (Supplemental Fig. 1).

2.6. Feature detection and dimensionality reduction based on the entire participants' sample

Feature detection was based on a univariate GLM analysis, intended to identify functional brain regions of interest (fROIs) that showed significant activation or deactivation in all four VSWM tasks, contrasted with all the fixation tasks, using the brain activity data of the ADHD boys and TD boys combined (total of 40 participants; see Chu et al., 2012; Friston, 2012 for discussions regarding optimal sample size). We found eight fROIs showing significant activation (VSWM > fixation) and eight fROIs showing significant deactivation (fixation > VSWM). We used a voxel threshold of $p < 0.01$ (family wise error [FWE] corrected), and voxel cluster threshold of $p < 0.01$ (minimum cluster size of 50 voxels; FWE), where each cluster had a single significant peak at $p < 0.05$ (FWE). An anatomical gray-matter mask (using the Talairach Daemon brain atlas gray-matter mask, with dilate = 3), and a sphere mask with a radius of 15 mm from each fROI peak voxel, constrained the fROI volume. These 16 ROIs likely reflect the VSWM network in a broad childhood population, in varying contexts (Fig. 2A).

For each participant we calculated the difference in mean Beta values between each of the four VSWM 2-back tasks and the mean of all fixation tasks in each of the 16 fROIs (2-back–fixation; fixation−2-back). Overall, given four VSWM tasks and 16 fROIs, the initial number of features, characterizing each participant, was 64 (see Supplemental Fig. 2 for activation profiles).

Given the relatively large initial number of features, and the high correlations between some of the features (see Supplemental Fig. 3 for the correlation matrix), we reduced the dimensionality of the data by using a sparse principal component analysis (SPCA). This resulted with a relatively small number of orthogonal principal components (PCs), which together explained most of the variability in the data, recalculated based only on the few features with the highest loadings (Berthet and Rigollet, 2013; Ritchie et al., 2015; Zou et al., 2006). This procedure has three major advantages: (i) SPCA enabled reducing the number of variables fed to the classifier to a number substantially smaller than the number of participants; (ii) recalculating the PCs based only on features with the highest loadings enabled excluding lower weight features that were likely to add mostly noise; and (iii) having each PC being affected by relatively few features, where each feature affects at most one PC, enabled better determining the underlying neurocognitive mechanisms represented by each PC.

We used sparse loading selection based on thresholding of the rotated loadings (loading threshold > 0.6; absolute weights > 0.1). Loading rotation was based on the Varimax rotation method (Lu and Zhang, 2012; Ma, 2013; Qi and Luo, 2015; Sjöstrand et al., 2006). The use of this threshold resulted with 39 features (out of 64) affecting the first 10 PCs (PCs with eigenvalues >2), where each feature affected a single PC (Fig. 2B). The first 10 PCs explained approximately 70% of the variability evident in the original 64 features. Each of the remaining PCs explained less than 3% of the variability in the data. Interestingly, we found each one of the first 10 PCs to reflect brain activity from several ROIs from the same VSWM task, but not brain activity in a specific ROI in several tasks. The exception was PC-2, which was based on the left and right middle frontal gyri in the two VSWM tasks without feedback. This supports our hypothesis that functional brain-imaging data from several distinct tasks is likely to add information useful for classification.

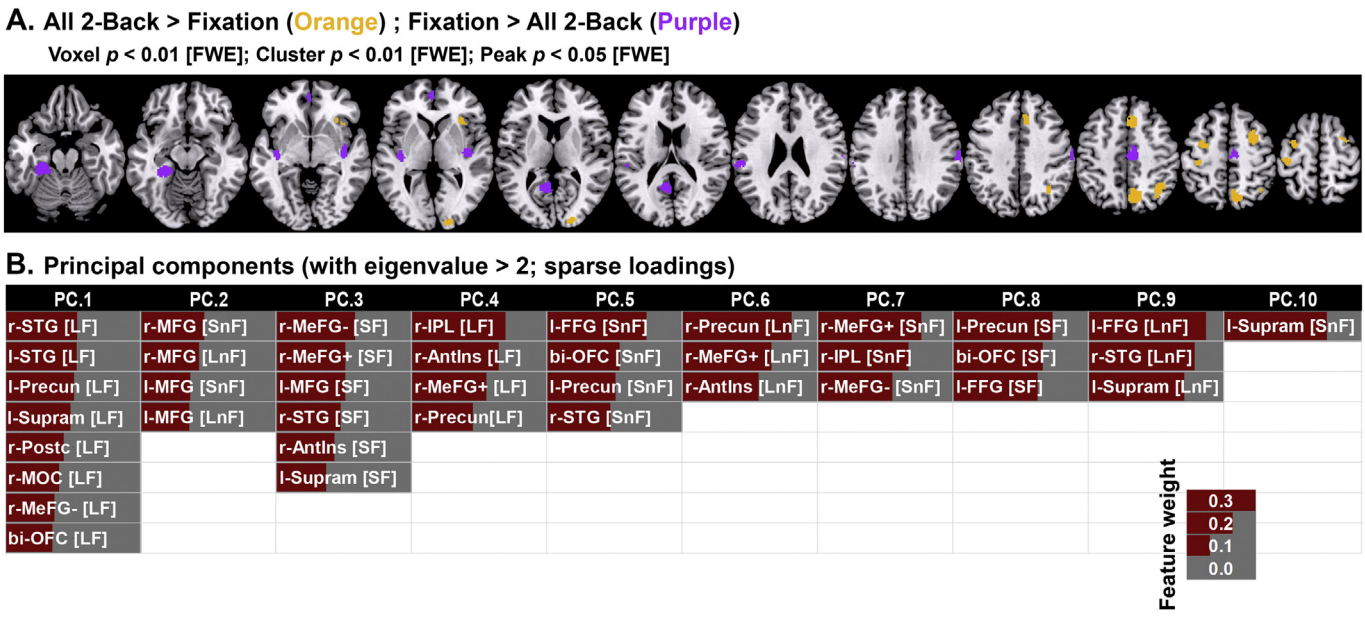

Fig. 2. Feature detection and dimensionality reduction. (A) Activated (orange) and deactivated fROIs (purple), in the VSWM tasks as contrasted with the fixation tasks. (B) The first 10 principal components' sparse loadings. The notation r-MeFG+ and r-MeFG− denote the activated and deactivated (respectively) right MeFG fROIs (see Supplemental Fig. 2). The experimental conditions are listed in brackets (LnF = large-reward, no-feedback; LF = large-reward, feedback; SnF = small-reward, no-feedback; SF = small-reward, feedback).

2.7. Logistic regression (LR) classification algorithm

For the learning of the ADHD classification model, we used a logistic regression (LR) classifier. The LR directly models the class conditional probabilities for each case, attempting to decisively classify as many cases as possible. A standard "jackknife" leave-one-out cross validation (Zion Golumbic et al., 2013; Ponce-Alvarez et al., 2012) was used for finding which of the 10 PCs significantly contributed to classification accuracy. In each cross validation iteration the LR classifier was trained based on the fMRI data of 39 participants, and then the goodness of fit of the learned model was evaluated based on the correlation between the clinician diagnosis of the left-out participant (ADHD or TD) and the LR classifier predicted ADHD probability for this participant. This procedure was repeated for all participants. The predictive power from all iterations was averaged to determine which of the 10 PCs are with statistically significant predictive power, and for estimating the predictive power of the final model. The reporting of a classification model based only on significant PCs has two advantages: (i) such a model is less likely to be overfitted or biased due to being based on too many predictors (PCs), and thus it provides a more conservative estimate of the classification accuracies that can be achieved with the data at hand and (ii) it further reduces the number of brain regions by which the two groups of interest (i.e., ADHD vs. TDs) differ, enabling a more parsimonious characterization of the differences in neurocognitive mechanisms between the two groups (see Supplemental Fig. 5 for an illustration of the data processing pipeline).

3. Results

3.1. ADHD classification using the fMRI data from the four tasks

Running the LR with the leave-one-out cross validation showed a statistically significant contribution for ADHD classification accuracies for only four PCs (PC-2, $p < 0.02$; PC-3, $p < 0.04$; PC-4, $p < 0.04$; PC-8, $p < 0.02$). Excluding each one of these four PCs from the classification model substantially impaired the classification accuracies, whereas adding any of the other six PCs did not increase the classification accuracies. The classification accuracy of the model based on the four statistically significant PCs was 92.5%, with 95% sensitivity and 90% specificity. Importantly, most of the classification decisions had high predicted probability values ($PP > 0.67$) assigned to most ADHD boys (75%), and low values ($PP < 0.33$) assigned to most TD boys (80%; Fig. 3), indicating a model with an excellent fit (omnibus test for model fit, $\chi^2(4) = 28.52$, $p < 0.0001$). This accuracy level is not statistically different from perfect accuracy reflecting the clinician's diagnosis (100%), $p = 0.12$ (one-tailed Fisher exact test; testing the hypothesis "LR classification is the same as the clinician diagnosis").

A permutation test shows that the classification accuracies of the above-described ADHD classification model are unlikely to be discovered by chance. In each permutation, labels of half the cases from each group were switched with the opposite group. The mean accuracy of 15 distinct permutations (62.0%) was significantly lower than that of the ADHD model (92.5%), $p = 0.000$ (exact test); $t(14) = 14.97$,

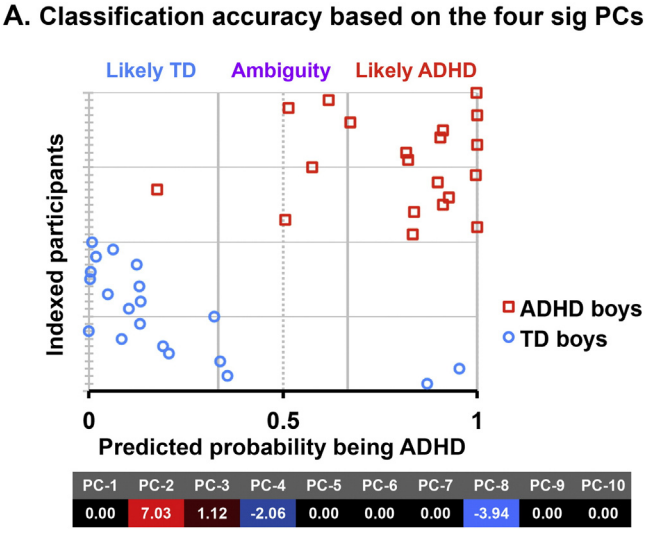

Fig. 3. ADHD classification accuracy as the predicted probability being an ADHD case. Color-labels represent the clinician's diagnosis for each case (ADHDs are red squares, and TDs are blue circles). The classifier decision boundary is at $PP = 0.5$. The $0.33 < PP < 0.67$ range represents an 'ambiguity zone' where classification decisions are made with lower confidence. Bottom — the respective weights (B coefficient) of each PC in the ADHD classification model. The weights of insignificant PCs are set to zero.

$p < 0.0001$ (one-tailed, one-sample t-test). Even the highest observed accuracy of a permuted-labels model was close to being significantly lower (15%) than that of the ADHD model, $p = 0.06$ (one-tailed Fisher exact test; testing the hypothesis "ADHD model is better than the best permuted model"). The mean correct decisive classification accuracies (with $PP < 0.33$ or $PP > 0.67$) in the 15 permutations were 24.5%, versus 77.5% of the ADHD model, $p = 0.000$ (exact test), $t(14) = 13.59$, $p < 0.0001$ (one-tailed, one-sample t-test; see also Supplemental Fig. 4).

3.2. ADHD classification based on each single VSWM task

Notably, the four significant PCs encompassed in the ADHD model are based on fMRI data from all four VSWM tasks: PC-2 is based on the two tasks without feedback; PC-3 is based on the small-reward with feedback task; PC-4 is based on the large-reward with feedback task; and PC-8 is based on the small-reward with feedback task (Fig. 2). In the following analysis, we further investigated the utility of using fMRI data from multiple VSWM tasks as compared with using fMRI data from a single task. Here we used the threshold of eigenvalue larger than one (keeping PCs that together explained more than 70% of the variance in the data).

ADHD classification accuracies based on the fMRI data from each single task (Fig. 4; Table 1) were significantly lower than the classification accuracy of the model based on all four tasks (Fig. 3). The exception was the classification accuracy based on the fMRI data from the small-reward with feedback task, which was 80%. However, here the correct decisive classification accuracy (with $PP < 0.33$ or $PP > 0.67$) based on the fMRI data from the small-reward with feedback VSWM task was 60%, which is significantly lower (17.5%) than the four-task classification accuracy, $p < 0.05$ (one-tailed Fisher exact test).

3.3. ADHD classification based on behavioral data

Feeding the LR classifier with the behavioral data from all four tasks resulted in an overall accuracy level of 75% (70% sensitivity; 80% specificity; omnibus test for model fit, $\chi^2(4) = 12.37$, $p < 0.01$). This overall accuracy level is significantly lower (17.5%) than the accuracy based on the fMRI data from the four tasks, $p < 0.04$ (one-tailed Fisher exact test; testing the hypothesis "fMRI model is better than the behavioral model"). Moreover, here only 57.5% of the classification decisions were both accurate and decisive, with high predicted probability values of being a boy with ADHD ($PP > 0.67$) assigned to only 50% of the ADHD boys, and low values ($PP < 0.33$) assigned to only 65% of the TD boys (Fig. 5B). This accuracy level is significantly lower (20%) than the decisive classification accuracy based on the four significant PCs (based on four tasks), $p < 0.04$ (one-tailed Fisher exact test).

Performance levels in the four tasks were highly correlated, where a boy that exhibited low performance in one task likely exhibited low performance in the other tasks. Moreover, we found high correlations between the behavioral performances in the four VSWM tasks and PC-2 (Fig. 5C). That is, the brain activity represented by PC-2 (right and left MFG, tasks without feedback) explains much of the observed variability in the behavioral data. On the other hand, the behavioral performances had low correlations with PC-3, PC-4 and PC-8, hence the additional diagnostic information provided by these PCs to the classification model. Feeding the LR classifier with both the behavioral and fMRI data (10 PCs) resulted in exactly the same model as when feeding it with only the fMRI data.

4. Discussion

4.1. Summary

We show that using a logistic regression (LR) classifier, which received as input brain imaging data that was acquired while children perform four distinct VSWM tasks, enabled an ADHD classification accuracy of 92.5% (Fig. 3). Classification accuracies that were based on the fMRI data from all four tasks were significantly higher than those obtained using the fMRI data from any single task. We found that the brain regions in which boys with ADHD exhibited altered pattern of brain activity differed from one task to the other. The corresponding behavioral data from the four VSWM tasks enabled an classification accuracy level of 75% (Fig. 5), which is significantly lower than those obtained using the fMRI data from the four tasks.

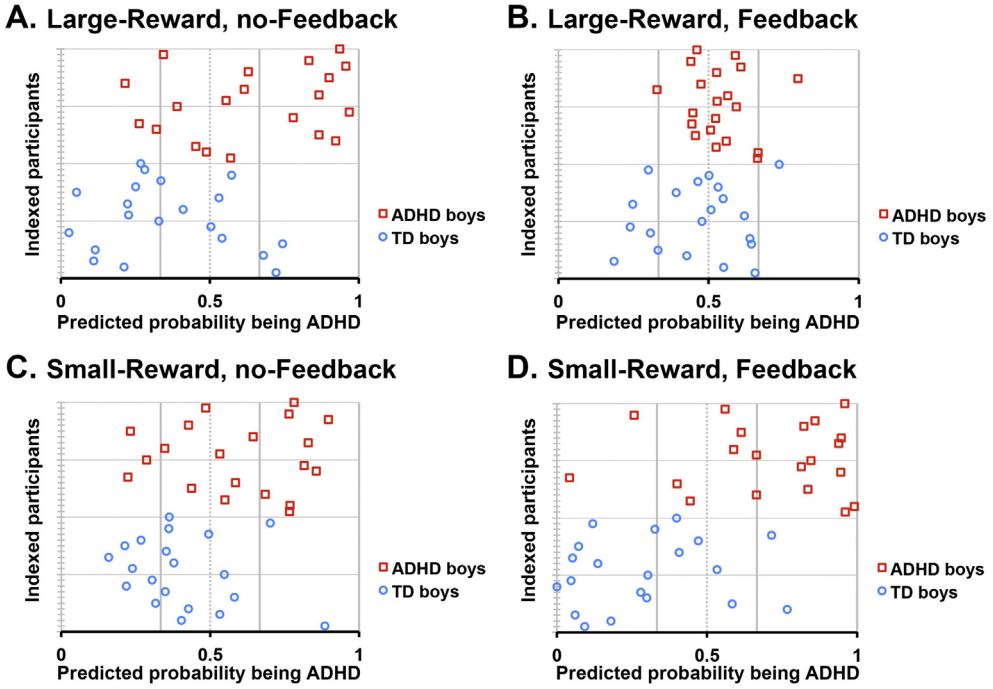

Fig. 4. Classification based on fMRI data from each single VSWM task. (A) Large-reward, no-feedback. (B) Large-reward, feedback. (C) Small-reward, no-feedback. (D) Small-reward, feedback. Classification accuracies based on the fMRI data from each single VSWM task were significantly lower than those obtained using the data from all four tasks (see Fig. 3 for comparison).

Table 1
Classification accuracies based on the fMRI from each individual VSWM task. Left to right column: Overall classification accuracy, sensitivity, specificity, omnibus test for model fit (and model significance), and the difference (in percentage) between the accuracy of the model based on a single task compared with the model based on the four tasks (significance is based on one-tailed Fisher exact test).

	Overall	Sensitivity	Specificity	Omnibus test	Difference
Large-reward No-feedback	65%	65%	65%	$\chi^2(1) = 13.70$ $p < 0.001$	27.5% $p < 0.003$
Large-reward Feedback	57.5%	65%	50%	$\chi^2(1) = 2.98$ $p = 0.084$	35% $p < 0.0003$
Small-reward No-feedback	70%	65%	75%	$\chi^2(1) = 7.96$ $p < 0.005$	22.5% $p < 0.01$
Small-reward Feedback	80%	80%	80%	$\chi^2(2) = 19.23$ $p < 0.001$	12.5% $p = 0.11$

Participants who were misclassified based on the data from one VSWM task were not necessarily misclassified when using the data from other VSWM tasks (Fig. 4). When the LR classifier was fed with the PCs calculated based on the four VSWM tasks together, we observed a substantial improvement in the classification accuracy (Fig. 3). This is consistent with the underlying theorem of ensemble-based classification (Rokach, 2010; Klöppel et al., 2012). As expected based on this theorem, classifying participants by using several measurements for each participant (e.g., fMRI data from few distinct VSWM tasks), where each measurement enables a better than chance classification accuracy, and where there is a substantial degree of independency between the measurements (e.g., distinct cognitive tasks), enabled higher classification accuracies compared to accuracies achieved by using the data from each single measurement.

4.2. Theoretical contribution

Our findings are with important theoretical contribution, providing additional support to the idea that the manifestation of neurocognitive abnormalities in ADHD is context dependent (Dovis et al., 2013; Hammer et al., 2015). Specifically, in the absence of trial-by-trial feedback, altered activity patterns characterizing ADHD were most evident in the bilateral middle frontal gyri (MFG), specifically the right-MFG (see PC-2 weight in Fig. 3; see feature weights in Fig. 2B). The MFG is believed to play an executive role in the visuospatial working memory network, and a primary role in volitional (top-down) allocation of attention (Burgess et al., 2010; Ehlis et al., 2008; Vance et al., 2007). We found that in VSWM tasks without feedback, boys with ADHD exhibited lower levels of activity in the MFG as compared with TD boys. This may indicate poor VSWM and poor capacity in allocating attention to target stimuli in children with ADHD, manifested when feedback is absent (see Cortese et al., 2012; Fassbender et al., 2011 for related findings).

The fMRI data from the small-reward with feedback VSWM task also had a considerable contribution to ADHD classification (PC-8 and PC-3; Fig. 3; Fig. 4B). Here, boys with ADHD exhibited altered brain activity in a network (PC-8) that primarily included the bilateral orbitofrontal cortex (bi-OFC) and the left fusiform gyrus (left-FFG). The OFC has been reported to play two primary roles that have potential relevance to the current task: Together with the inferior frontal gyrus, the anterior insula, the superior temporal cortex and the temporoparietal junction, the OFC is part of the ventral attention network, acting as a bottom-up saliency detection system determining subjective and context-dependent susceptibility to unexpected salient stimuli (Corbetta et al., 2008; Vossel et al., 2014; Weissman and Prado, 2012). The OFC was also found to be involved in the processing of reward-related information (Schoenbaum and Roesch, 2005; Pauli et al., 2012). It is suggested that the OFC is involved in monitoring which recent actions were rewarded, and predicting which future actions are most likely to be rewarded (Kahnt et al., 2010). The left-FFG was found to be involved in letter identification and reading (McCandliss et al., 2003; Dehaene and Cohen, 2011; McNorgan et al., 2013). Here we found greater deactivation in these two brain regions in TD boys, as compared to boys with ADHD, primarily in the small-reward with feedback VSWM task. This may indicate that boys with ADHD fail in suppressing task irrelevant information (letter identity in a task that requires monitoring the spatial location of letters, and visual feedback indicating insignificant reward).

The features with significant loadings on PC-3 (small-reward with feedback) were the two fROIs in the right medial/superior prefrontal gyrus (right-MeFG- and right-MeFG+, Fig. 2B), the left-MFG, the right superior temporal gyrus (right-STG), the right anterior insula (right-AntIns), and the right supramarginal gyrus. These brain regions are involved in feedback processing (Ferdinand and Opitz, 2014), bottom-up attention and in sensory integration (Prado et al., 2011; Weissman and Prado, 2012). The loadings on PC-3 (and PC-8), indicate that in the small-reward with feedback condition children with ADHD are likely to exhibit altered suppression of irrelevant visual features

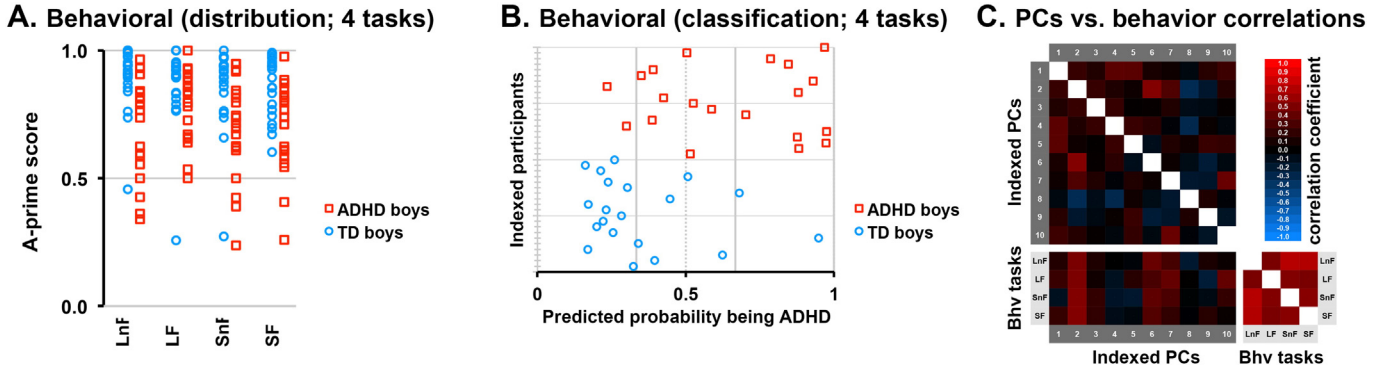

Fig. 5. Behavioral data. (A) Performance distributions in the four experimental tasks. LnF = large-reward, no-feedback; LF = large-reward, feedback; SnF = small-reward, no-feedback; SF = small-reward, feedback (see also Supplemental Table 3). (B) Classification based on the behavioral data from all four VSWM tasks. (C) Correlations between the 10 PCs, and between the behavioral performances in the four tasks (see online higher resolution version of this figure).

(FFG), altered visuospatial processing (precuneus), altered sensory integration (STG and supramarginal), altered bottom-up attention control (OFC), altered feedback processing and top-down attention control (MeFG and MFG), and altered synchronization between bottom-up and top-down attention control (AntIns).

The fMRI data from the large-reward with feedback VSWM task had the smallest contribution to ADHD classification (57.5% accuracy, with only a trend toward a significant model fit; Fig. 4B). Altered activity in ADHD in this VSWM task was evident in PC-4 (Fig. 3), which included the right inferior parietal lobe (right-IPL), the right-AntIns, the right-MeFG (the activated right-MeFG fROI; see Fig. 2) and the right precuneus. These fROIs partially overlap (right-AntIns and right-MeFG) with those included in PC-3 (small-reward with feedback). This implies that the right-AntIns and right-MeFG are associated with altered feedback processing in ADHD (regardless of reward expectation). This is consistent with earlier findings showing that the right-AntIns and the right-MeFG (together with the anterior cingulate) mediate between the central executive network and brain regions involved in risk/gain prediction, where response selection is required (Menon and Udin, 2010; Preuschoff et al., 2008; Späti et al., 2014; Taylor et al., 2009). Early studies showed that children with ADHD exhibit poor cognitive control associated with lower levels of neural activity in the anterior insula, as compared with TDs (Cubillo et al., 2010; Morein-Zamir et al., 2014).

4.3. Limitations and future research

We show that using fMRI data from four distinct VSWM tasks enabled substantially better classification accuracies than the use of fMRI data from a single VSWM task. Nevertheless, the current investigation was limited to the manipulation of feedback and reward in VSWM tasks, where many characteristics of the four tasks were identical. It is possible that an even better ADHD classification can be achieved by using other, perhaps more demanding cognitive tasks that require other cognitive skills impaired in ADHD (e.g., tasks with specifically greater response inhibition demands; Booth et al., 2005; Rubia et al., 2005). Moreover, the duration of each VSWM task we used here was 48 trials long (~100 s). It is possible that these tasks can be shortened (e.g., reduced to 32, or even fewer trials) without compromising classification accuracies. As part of the development of a practical diagnosis tool, future investigations should aim to identify an optimized ensemble of cognitive tasks that would yield the highest classification accuracies in the shortest scanning time possible. This would enable reducing scanning costs and potential discomfort to the scanned subjects. Future investigation may also involve looking for an ensemble of tasks that enable differentiation between children with ADHD from other clinical populations, with shared symptoms. The use of multi-task-based fMRI data may also enable earlier diagnosis of ADHD, prior to the onset of clear behavioral symptoms, which in turn may enable earlier intervention.

In order to be of practical clinical use, task-based fMRI diagnosis requires the scanned participants to accurately perform a cognitive task while keeping still throughout a relatively long scan. Thus, a multi-task-based fMRI diagnosis session may not be practical for very young children or individuals with severe cognitive deficits. However, mild ADHD and moderate ADHD are much more common and represents more of a diagnostic dilemma for clinicians. In contrast, severe or very early onset ADHD would likely have a more distinct etiology. Future developments should ideally involve an integrative use of multi-task-based fMRI, resting-state fMRI and structural imaging. This would likely enable an effective diagnosis of most clinical cases, and the detection of the onset of some clinical condition in early childhood. Ideally, future imaging-based diagnostic tools may enable finer differentiation of ADHD subgroups, assisting clinicians in customizing intervention.

It is likely that similarly high (or even higher) classification accuracies can be achieved by using other machine-learning based classification methods. Future studies should specifically explore machine-learning methods based on whole brain data, from multiple tasks, which automatically detect brain regions in which the targeted clinical population exhibits abnormal patterns of brain activity (e.g., Ryali et al., 2010).

4.4. Conclusions

We show that fMRI data acquired while participants perform a few distinct cognitive tasks enables substantial improvement in the detection of ADHD cases, as compared with the use of fMRI data from a single task (or corresponding behavioral data). Showing that fMRI data enables better classification accuracies than the corresponding behavioral data suggests that even ADHD cases that exhibited normal-like VSWM performance were likely to be characterized by substantially altered pattern of brain activation. This provides a proof-of-concept that scanning subjects while they perform an ensemble of distinct cognitive tasks is likely to payoff, enabling greater accuracies and higher confidence diagnosis of clinical cases. We suggest that this approach can be used for diagnosing other clinical populations, and possibly also for dissociating between distinct clinical populations who share behavioral symptoms. This would require using cognitive tasks that target the neurocognitive deficits characterizing the clinical condition of interest, or a battery of tasks that may enable dissociating between distinct neurocognitive abnormalities.

Acknowledgments

This study was funded by the NIH grant R21-MH080820 to J. R. Booth, and the Northwestern University Human Cognition T32-NS047987 NIH training grant to R. Hammer. Parts of the reported methods and findings are included in provisional patent, NU2014-051-01 US. We thank Jessica Gayda, Michael Tennekoon, Shira DeCovnick, Audrey Haque, Jessica Stroup and the other members of the Developmental Cognitive Neuroscience Lab, at Northwestern University, for their help in data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.nicl.2015.08.015.

References

ADHD-200 Consortium, 2012. A model to advance the translational potential of neuroimaging in clinical neuroscience. Front. Syst. Neurosci. 6.

Anderson, A., Douglas, P.K., Kerr, W.T., Haynes, V.S., Yuille, A.L., Xie, J., Cohen, M.S., 2014. Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. NeuroImage 102, 207–219. http://dx.doi.org/10.1016/j.neuroimage.2013.12.015.

Asherson, P., Akehurst, R., Kooij, J.J.S., Huss, M., Beusterien, K., Sasané, R., Hodgkins, P., 2012. Under diagnosis of adult ADHD: cultural influences and societal burden. J. Atten. Disord. 16 (5), 20S–38S. http://dx.doi.org/10.1177/1087054711435360.

Berthet, Q., Rigollet, P., 2013. Optimal detection of sparse principal components in high dimension. Ann. Statist. 41 (4), 1780–1815. http://dx.doi.org/10.1214/13-AOS1127.

Booth, J.R., Burman, D.D., Meyer, J.R., Lei, Z., Trommer, B.L., Davenport, N.D., Li, W., Parrish, T.B., Gitelman, D.R., Mesulam, M.M., 2005. Larger deficits in brain networks for response inhibition than for visual selective attention in attention deficit hyperactivity disorder (ADHD). J. Child Psychol. Psychiatry 46 (1), 94–111. http://dx.doi.org/10.1111/j.1469-7610.2004.00337.x15660647.

Boyle, C.A., Boulet, S., Schieve, L.A., Cohen, R.A., Blumberg, S.J., Yeargin-Allsopp, M., Visser, S., Kogan, M.D., 2011. Trends in the prevalence of developmental disabilities in US children, 1997–2008. Pediatrics 127 (6), 1034–1042. http://dx.doi.org/10.1542/peds.2010-298921606152.

Burgess, G.C., Depue, B.E., Ruzic, L., Willcutt, E.G., Du, Y.P., Banich, M.T., 2010. Attentional control activation relates to working memory in attention-deficit/hyperactivity disorder. Biol. Psychiatry 67 (7), 632–640. http://dx.doi.org/10.1016/j.biopsych.2009.10.03620060961.

Burgund, E.D., Kang, H.C., Kelly, J.E., Buckner, R.L., Snyder, A.Z., Petersen, S.E., Schlaggar, B.L., 2002. The feasibility of a common stereotactic space for children and adults in fMRI studies of development. Neuroimage 17 (1), 184–200. http://dx.doi.org/10.1006/nimg.2002.117412482076.

Calhoun, V.D., Adali, T., Stevens, M.C., Kiehl, K.A., Pekar, J.J., 2005. Semi-blind ICA of fMRI: a method for utilizing hypothesis-derived time courses in a spatial ICA analysis. Neuroimage 25 (2), 527–538. http://dx.doi.org/10.1016/j.neuroimage.2004.12.012.

Castellanos, F.X., Proal, E., 2012. Large-scale brain systems in ADHD: beyond the prefrontal–striatal model. Trends Cogn. Sci. (Regul. Ed.) 16 (1), 17–26. http://dx.doi.org/10.1016/j.tics.2011.11.00722169776.

Chu, C., Hsu, A.L., Chou, K.H., Bandettini, P., Lin, C., 2012. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage 60 (1), 59–70. http://dx.doi.org/10.1016/j.neuroimage.2011.11.066.

Clark, L., Blackwell, A.D., Aron, A.R., Turner, D.C., Dowson, J., Robbins, T.W., Sahakian, B.J., 2007. Association between response inhibition and working memory in adult ADHD: a link to right frontal cortex pathology? Biol. Psychiatry 61 (12), 1395–1401. http://dx.doi.org/10.1016/j.biopsych.2006.07.02017046725.

Corbetta, M., Patel, G., Shulman, G.L., 2008. The reorienting system of the human brain: from environment to theory of mind. Neuron 58 (3), 306–324. http://dx.doi.org/10.1016/j.neuron.2008.04.01718466742.

Cortese, S., Kelly, C., Chabernaud, C., Proal, E., Di Martino, A., Milham, M.P., Castellanos, F.X., 2012. Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. Am. J. Psychiatry 169 (10), 1038–1055. http://dx.doi.org/10.1176/appi.ajp.2012.1110152122983386.

Cubillo, A., Halari, R., Ecker, C., Giampietro, V., Taylor, E., Rubia, K., 2010. Reduced activation and inter-regional functional connectivity of fronto-striatal networks in adults with childhood attention-deficit hyperactivity disorder (ADHD) and persisting symptoms during tasks of motor inhibition and cognitive switching. J. Psychiatr. Res. 44 (10), 629–639. http://dx.doi.org/10.1016/j.jpsychires.2009.11.016.

Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. Trends Cogn. Sci. 15 (6), 254–262. http://dx.doi.org/10.1016/j.tics.2011.04.003.

del Campo, N., Fryer, T.D., Hong, Y.T., Smith, R., Brichard, L., Acosta-Cabronero, J., Müller, U., 2013. A positron emission tomography study of nigro-striatal dopaminergic mechanisms underlying attention: implications for ADHD and its treatment. Brain 136 (11), 3252–3270. http://dx.doi.org/10.1093/brain/awt263.

Demir, Ö.E., Prado, J., Booth, J.R., 2014. The differential role of verbal and spatial working memory in the neural basis of arithmetic. Dev. Neuropsychol. 39 (6), 440–458. http://dx.doi.org/10.1080/87565641.2014.939182.

Demurie, E., Roeyers, H., Baeyens, D., Sonuga-Barke, E., 2011. Common alterations in sensitivity to type but not amount of reward in ADHD and autism spectrum disorders. J. Child Psychol. Psychiatry 52 (11), 1164–1173. http://dx.doi.org/10.1111/j.1469-7610.2010.02374.x.

Dovis, S., Van der Oord, S., Huizenga, H.M., Wiers, R.W., Prins, P.J.M., 2015. Prevalence and diagnostic validity of motivational impairments and deficits in visuospatial short-term memory and working memory in ADHD subtypes. Eur. Child Adolesc. Psychiatry 24 (5), 575–590. http://dx.doi.org/10.1007/s00787-014-0612-1.

Dovis, S., Van der Oord, S., Wiers, R.W., Prins, P.J.M., 2013. What part of working memory is not working in ADHD? Short-term memory, the central executive and effects of reinforcement. J. Abnorm. Child Psychol. 41 (6), 901–917. http://dx.doi.org/10.1007/s10802-013-9729-9.

DuPaul, G.J., Power, T.J., Anastopoulos, A.D., Reid, R., 1998. ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation 25. Guilford Press, New York.

Ehlis, A.C., Bähne, C.G., Jacob, C.P., Herrmann, M.J., Fallgatter, A.J., 2008. Reduced lateral prefrontal activation in adult patients with attention-deficit/hyperactivity disorder ADHD) during a working memory task: a functional near-infrared spectroscopy (fNIRS) study. J. Psychiatr. Res. 42 (13), 1060–1067. http://dx.doi.org/10.1016/j.jpsychires.2007.11.011.

Fassbender, C., Schweitzer, J.B., Cortes, C.R., Tagamets, M.A., Windsor, T.A., Reeves, G.M., Gullapalli, R., 2011. Working memory in attention deficit/hyperactivity disorder is characterized by a lack of specialization of brain function. PLOS ONE 6 (11), e27240. http://dx.doi.org/10.1371/journal.pone.0027240.

Ferdinand, N.K., Opitz, B., 2014. Different aspects of performance feedback engage different brain areas: disentangling valence and expectancy in feedback processing. Sci. Rep. 4, 5986. http://dx.doi.org/10.1038/srep05986.

Friston, K., 2012. Ten ironic rules for non-statistical reviewers. Neuroimage 61 (4), 1300–1310. http://dx.doi.org/10.1016/j.neuroimage.2012.04.018.

Getahun, D., Jacobsen, S.J., Fassett, M.J., Chen, W., Demissie, K., Rhoads, G.G., 2013. Recent trends in childhood attention-deficit/hyperactivity disorder. J.A.M.A. Pediatr. 167 (3), 282–288. http://dx.doi.org/10.1001/2013.jamapediatrics.401.

Ghosh, S.S., Kakunoori, S., Augustinack, J., Nieto-Castanon, A., Kovelman, I., Gaab, N., Fischl, B., 2010. Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11years of age. Neuroimage 53 (1), 85–93. http://dx.doi.org/10.1016/j.neuroimage.2010.05.075.

Hale, T.S., Wiley, J.F., Smalley, S.L., Tung, K.L., Kaminsky, O., McGough, J.J., Loo, S.K., 2015. A parietal biomarker for ADHD liability: as predicted by the distributed effects perspective model of ADHD. Front. Psychiatry 6. http://dx.doi.org/10.3389/fpsyt.2015.00063.

Hammer, R., Tennekoon, M., Cooke, G.E., Gayda, J., Stein, M.A., Booth, J.R., 2015. Feedback associated with expectation for larger-reward improves visuospatial working memory performances in children with ADHD. Dev. Cogn. Neurosci. 14, 38–49. http://dx.doi.org/10.1016/j.dcn.2015.06.002.

Hart, H., Chantiluke, K., Cubillo, A.I., Smith, A.B., Simmons, A., Brammer, M.J., Rubia, K., 2014a. Pattern classification of response inhibition in ADHD: toward the development of neurobiological markers for ADHD. Hum. Brain Mapp. 35 (7), 3083–3094. http://dx.doi.org/10.1002/hbm.22386.

Hart, H., Marquand, A.F., Smith, A., Cubillo, A., Simmons, A., Brammer, M., Rubia, K., 2014b. Predictive neurofunctional markers of attention-deficit/hyperactivity disorder based on pattern classification of temporal processing. J. Am. Acad. Child Adolesc. Psychiatry 53 (5), 569–578.e1. http://dx.doi.org/10.1016/j.jaac.2013.12.024.

Hoekzema, E., Carmona, S., Ramos-Quiroga, J.A., Richarte Fernández, V., Bosch, R., Soliva, J.C., ... Vilarroya, O., 2014. An independent components and functional connectivity analysis of resting state fMRI data points to neural network dysregulation in adult ADHD. Hum. Brain Mapp. 35 (4), 1261–1272. http://dx.doi.org/10.1002/hbm.22250.

Iannaccone, R., Hauser, T.U., Ball, J., Brandeis, D., Walitza, S., Brem, S., 2015. Classifying adolescent attention-deficit/hyperactivity disorder (ADHD) based on functional and structural imaging. Eur. Child Adolesc. Psychiatry 1–11 http://dx.doi.org/10.1007/s00787-015-0678-4.

Johnston, B.A., Mwangi, B., Matthews, K., Coghill, D., Konrad, K., Steele, J.D., 2014. Brainstem abnormalities in attention deficit hyperactivity disorder support high accuracy individual diagnostic classification. Hum. Brain Mapp. 35 (10), 5179–5189. http://dx.doi.org/10.1002/hbm.22542.

Kahnt, T., Heinzle, J., Park, S.Q., Haynes, J.-D., 2010. The neural code of reward anticipation in human orbitofrontal cortex. Proc. Natl. Acad. Sci. 107 (13), 6010–6015. http://dx.doi.org/10.1073/pnas.0912838107.

Kaufman, J., Birmaher, B., Brent, D., Rao, U.M.A., Flynn, C., Moreci, P., Ryan, N., 1997. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J. Am. Acad. Child Adolesc. Psychiatry 36 (7), 980–988. http://dx.doi.org/10.1097/00004583-199707000-00021.

Klöppel, S., Abdulkadir, A., Jack, C.R., Koutsouleris, N., Mourão-Miranda, J., Vemuri, P., 2012. Diagnostic neuroimaging across diseases. Neuroimage 61 (2), 457–463. http://dx.doi.org/10.1016/j.neuroimage.2011.11.002.

Larson, K., Russ, S.A., Kahn, R.S., Halfon, N., 2011. Pattern of comorbidity, functioning, and service use for US children with ADHD, 2007. Pediatrics 127 (3), 462–470.

Lim, L., Marquand, A., Cubillo, A.A., Smith, A.B., Chantiluke, K., Simmons, A., Rubia, K., 2013. Disorder-specific predictive classification of adolescents with attention deficit hyperactivity disorder (ADHD) relative to autism using structural magnetic resonance imaging. PLOS ONE 8 (5), e63660. http://dx.doi.org/10.1371/journal.pone.0063660.

Lu, Z., Zhang, Y., 2012. An augmented Lagrangian approach for sparse principal component analysis. Math. Program. 135 (1–2), 149–193. http://dx.doi.org/10.1007/s10107-011-0452-4.

Ma, Z., 2013. Sparse principal component analysis and iterative thresholding. Ann. Statist. 41 (2), 772–801. http://dx.doi.org/10.1214/13-AOS1097.

McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn. Sci. 7 (7), 293–299. http://dx.doi.org/10.1016/S1364-6613(03)00134-7.

McNorgan, C., Randazzo-Wagner, M., Booth, J.R., 2013. Cross-modal integration in the brain is related to phonological awareness only in typical readers, not in those with reading difficulty. Front. Hum. Neurosci. 7. http://dx.doi.org/10.3389/fnhum.2013.00388.

Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. 214 (5–6), 655–667. http://dx.doi.org/10.1007/s00429-010-0262-0.

Morein-Zamir, S., Dodds, C., van Hartevelt, T.J., Schwarzkopf, W., Sahakian, B., Müller, U., Robbins, T., 2014. Hypoactivation in right inferior frontal cortex is specifically associated with motor response inhibition in adult ADHD. Hum. Brain Mapp. 35 (10), 5141–5152. http://dx.doi.org/10.1002/hbm.22539.

Morgan, P.L., Staff, J., Hillemeier, M.M., Farkas, G., Maczuga, S., 2013. Racial and ethnic disparities in ADHD diagnosis from kindergarten to eighth grade. Pediatrics 132 (1), 85–93. http://dx.doi.org/10.1542/peds.2012-2390.

Morris, R.W., Sparks, A., Mitchell, P.B., Weickert, C.S., Green, M.J., 2012. Lack of cortico-limbic coupling in bipolar disorder and schizophrenia during emotion regulation. Transl. Psychiatry 2 (3), e90. http://dx.doi.org/10.1038/tp.2012.16.

Mulligan, R.C., Knopik, V.S., Sweet, L.H., Fischer, M., Seidenberg, M., Rao, S.M., 2011. Neural correlates of inhibitory control in adult attention deficit/hyperactivity disorder: evidence from the Milwaukee longitudinal sample. Psychiatry Res. Neuroimag. 194 (2), 119–129. http://dx.doi.org/10.1016/j.pscychresns.2011.02.003.

Nouretdinov, I., Costafreda, S.G., Gammerman, A., Chervonenkis, A., Vovk, V., Vapnik, V., Fu, C.H.Y., 2011. Machine learning classification with confidence: application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. Neuroimage 56 (2), 809–813. http://dx.doi.org/10.1016/j.neuroimage.2010.05.023.

Pauli, W.M., Hazy, T.E., O'Reilly, R.C., 2012. Expectancy, ambiguity, and behavioral flexibility: separable and complementary roles of the orbital frontal cortex and amygdala in processing reward expectancies. J. Cogn. Neurosci. 24 (2), 351–366. http://dx.doi.org/10.1162/jocn_a_00155.

Peng, X., Lin, P., Zhang, T., Wang, J., 2013. Extreme learning machine-based classification of ADHD using brain structural MRI data. PLOS ONE 8 (11), e79476. http://dx.doi.org/10.1371/journal.pone.0079476.

Peters, S., Koolschijn, P.C.M.P., Crone, E.A., Van Duijvenvoorde, A.C.K., Raijmakers, M.E.J., 2014. Strategies influence neural activity for feedback learning across child and adolescent development. Neuropsychologia 62, 365–374. http://dx.doi.org/10.1016/j.neuropsychologia.2014.07.006.

Plichta, M.M., Scheres, A., 2014. Ventral–striatal responsiveness during reward anticipation in ADHD and its relation to trait impulsivity in the healthy population: a meta-analytic review of the fMRI literature. Neurosci. Biobehav. Rev. 38, 125–134. http://dx.doi.org/10.1016/j.neubiorev.2013.07.012.

Plichta, M.M., Vasic, N., Wolf, R.C., Lesch, K.P., Brummer, D., Jacob, C., ... Grön, G., 2009. Neural hyporesponsiveness and hyperresponsiveness during immediate and delayed reward processing in adult attention-deficit/hyperactivity disorder. Biological psychiatry 65 (1), 7–14.

Polanczyk, G.V., Willcutt, E.G., Salum, G.A., Kieling, C., Rohde, L.A., 2014. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. Int. J. Epidemiol. 43 (2), 434–442. http://dx.doi.org/10.1093/ije/dyt261.

Ponce-Alvarez, A., Nácher, V., Luna, R., Riehle, A., Romo, R., 2012. Dynamics of cortical neuronal ensembles transit from decision making to storage for later report. J. Neurosci. 32 (35), 11956–11969. http://dx.doi.org/10.1523/JNEUROSCI.6176-11.2012.

Prado, J., Carp, J., Weissman, D.H., 2011. Variations of response time in a selective attention task are linked to variations of functional connectivity in the attentional network. Neuroimage 54 (1), 541–549. http://dx.doi.org/10.1016/j.neuroimage.2010.08.022.

Preuschoff, K., Quartz, S.R., Bossaerts, P., 2008. Human insula activation reflects risk prediction errors as well as risk. J. Neurosci. 28 (11), 2745–2752. http://dx.doi.org/10.1523/JNEUROSCI.4286-07.2008.

Puschmann, S., Brechmann, A., Thiel, C.M., 2013. Learning-dependent plasticity in human auditory cortex during appetitive operant conditioning. Hum. Brain Mapp. 34 (11), 2841–2851. http://dx.doi.org/10.1002/hbm.22107.

Qi, X., Luo, R., 2015. Sparse principal component analysis in Hilbert space. Scand. J. Stat. 42 (1), 270–289. http://dx.doi.org/10.1111/sjos.12106.

Ritchie, M.D., Holzinger, E.R., Li, R., Pendergrass, S.A., Kim, D., 2015. Methods of integrating data to uncover genotype–phenotype interactions. Nat. Rev. Genet. 16 (2), 85–97. http://dx.doi.org/10.1038/nrg3868.

Rokach, L., 2010. Ensemble-based classifiers. Artif. Intell. Rev. 33 (1–2), 1–39. http://dx.doi.org/10.1007/s10462-009-9124-7.

Rubia, K., Smith, A.B., Brammer, M.J., Toone, B., Taylor, E., 2005. Abnormal brain activation during inhibition and error detection in medication-naive adolescents with ADHD. Am. J. Psychiatry 162 (6), 1067–1075. http://dx.doi.org/10.1176/appi.ajp.162.6.1067.

Ryali, S., Supekar, K., Abrams, D.A., Menon, V., 2010. Sparse logistic regression for whole-brain classification of fMRI data. NeuroImage 51 (2), 752–764. http://dx.doi.org/10.1016/j.neuroimage.2010.02.040.

Schoenbaum, G., Roesch, M., 2005. Orbitofrontal cortex, associative learning, and expectancies. Neuron 47 (5), 633–636. http://dx.doi.org/10.1016/j.neuron.2005.07.018.

Sjöstrand, K., Stegmann, M.B., Larsen, R., 2006. Sparse principal component analysis in medical shape modeling. J. Med. Imaging 26 (12), 1625–1635.

Sonuga-Barke, E.J.S., Fairchild, G., 2012. Neuroeconomics of attention-deficit/hyperactivity disorder: differential influences of medial, dorsal, and ventral prefrontal brain networks on suboptimal decision making? Biol. Psychiatry 72 (2), 126–133. http://dx.doi.org/10.1016/j.biopsych.2012.04.004.

Späti, J., Chumbley, J., Brakowski, J., Dörig, N., Grosse Holtforth, M., Seifritz, E., Spinelli, S., 2014. Functional lateralization of the anterior insula during feedback processing. Hum. Brain Mapp. 35 (9), 4428–4439. http://dx.doi.org/10.1002/hbm.22484.

Taylor, K.S., Seminowicz, D.A., Davis, K.D., 2009. Two systems of resting state connectivity between the insula and cingulate cortex. Hum. Brain Mapp. 30 (9), 2731–2745. http://dx.doi.org/10.1002/hbm.20705.

Thoma, V., Henson, R.N., 2011. Object representations in ventral and dorsal visual streams: fMRI repetition effects depend on attention and part–whole configuration. Neuroimage 57 (2), 513–525. http://dx.doi.org/10.1016/j.neuroimage.2011.04.035.

Tomasi, D., Volkow, N.D., 2014. Functional connectivity of substantia nigra and ventral tegmental area: maturation during adolescence and effects of ADHD. Cereb. Cortex 24 (4), 935–944. http://dx.doi.org/10.1093/cercor/bhs382.

van der Schaaf, M.E., Fallon, S.J., ter Huurne, N., Buitelaar, J., Cools, R., 2013. Working memory capacity predicts effects of methylphenidate on reversal learning. Neuropsychopharmacology 38 (10), 2011–2018. http://dx.doi.org/10.1038/npp.2013.100.

Vance, A., Silk, T.J., Casey, M., Rinehart, N.J., Bradshaw, J.L., Bellgrove, M.A., Cunnington, R., 2007. Right parietal dysfunction in children with attention deficit hyperactivity disorder, combined type: a functional MRI study. Mol. Psychiatry 12 (9), 826–832. http://dx.doi.org/10.1038/sj.mp.4001999.

Volkow, N.D., Wang, G.J., Kollins, S.H., Wigal, T.L., Newcorn, J.H., Telang, F., Swanson, J.M., 2009. Evaluating dopamine reward pathway in ADHD. JAMA 302 (10), 1084–1091. http://dx.doi.org/10.1001/jama.2009.1308.

Vossel, S., Geng, J.J., Fink, G.R., 2014. Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. The Neuroscientist 20 (2), 150–159. http://dx.doi.org/10.1177/1073858413494269.

Weissman, D.H., Prado, J., 2012. Heightened activity in a key region of the ventral attention network is linked to reduced activity in a key region of the dorsal attention network during unexpected shifts of covert visual spatial attention. Neuroimage 61 (4), 798–804. http://dx.doi.org/10.1016/j.neuroimage.2012.03.032.

Zhang, Y., Inder, T.E., Neil, J.J., Dierker, D.L., Alexopoulos, D., Anderson, P.J., Van Essen, D.C., 2015. Cortical structural abnormalities in very preterm children at 7 years of age. NeuroImage 109 (1), 469–479.

Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron 77 (5), 980–991. http://dx.doi.org/10.1016/j.neuron.2012.12.037.

Zou, H., Hastie, T., Tibshirani, R., 2006. Sparse principal component analysis. J. Comput. Graph. Stat. 15 (2), 265–286. http://dx.doi.org/10.1198/106186006X113430.