Supporting Information

Fair et al. 10.1073/pnas.1115365109

SI Text

Subjects and Demographics. Children ages 6–17 y participated in this study. They were recruited via widespread community outreach (public advertisements, mass mailings to advertiser lists, radio ads, and fliers at clinics) to recruit a sample that would not be biased by clinical referral but would reflect children with possible attention deficit/hyperactivity disorder (ADHD) and typically developing controls (TDC). Children were evaluated with an extensive battery of measures including a structured clinical interview, parent and teacher rating scales, and intelligence quotient (IQ) and achievement measures. These data were combined to create a best estimate assignment of children to either ADHD or TDC groups (children who were not able to be reliably clinically assigned due to borderline symptom scores were excluded from the present analysis), as detailed elsewhere (1). Demographic details are listed in Table S1.

Diagnostic Evaluation. Psychiatric diagnoses were evaluated with the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS-I) (2) administered to a parent and with a parent and teacher *Conners Rating Scale*, Third Edition (3). Intelligence was evaluated with a three-subtest short form (block design, vocabulary, and information) of the *Wechsler Intelligence Scale for Children*, Fourth Edition (WISC-IV) (4). A diagnostic team (a board-certified child psychiatrist and licensed clinical psychologist) independently reviewed the case records and interviewer notes to arrive at a decision regarding ADHD (and ADHD subtype) and comorbid disorders. If they disagreed, the case was conferenced; if consensus was not easily obtained, the case was excluded. Their agreement rates were acceptable ($\kappa > 0.80$ for all disorders with base rate >5%). Children were classified as ADHD combined subtype if they currently met criteria for ADHD and ever met criteria for combined subtype and as primarily inattentive subtype if they met criteria for ADHD and always met criteria for inattentive subtype. We excluded two children with the primarily hyperactive subtype from this report. For purposes of the current study, all children with ADHD were pooled into a single group.

Children were excluded if they did not clearly meet criteria for ADHD or non-ADHD groups (i.e., children deemed subthreshold by the clinicians were excluded). Children were also excluded if a history of neurological illness, chronic medical problems, sensorimotor handicap, autistic disorder, mental retardation, or significant head trauma (with loss of consciousness) was identified by parent report or if they had evidence of psychotic disorder or bipolar disorder on the structured parent psychiatric interview. Children currently prescribed nonstimulant psychotropic medications (including atomoxetine) were excluded. Children prescribed short-acting stimulant medications underwent neuropsychological testing after a minimum five half-life washout (i.e., 24–48 h depending on the preparation). Typically developing control children were excluded for presence of conduct disorder, major depressive disorder, or history of psychotic disorder, as well as for presence of ADHD.

Background Measures of Cognitive Functioning. Youth completed two laboratory sessions. The first session was composed of the diagnostic screening along with other testing, which included a short form of the WISC-IV (4). The age-adjusted standardized score was used as the estimate of full-scale IQ for each child. They also completed the *Wechsler Individual Achievement Test*, Second Edition (WIAT-II) (5) word reading subtest so we could screen for possible learning disorder. The second session included the experimental battery as described below.

Neuropsychological Measures Theorized to Relate to ADHD. We devised a battery of measures intended to capture many different hypotheses about possible cognitive problems in ADHD. The battery thus was designed to capture working memory (6), response inhibition (7, 8), response variability (9), temporal information processing (10), arousal and activation (11), interference control (12), and response speed (13). Because our focus was on cognitive types and not motivation or emotion/reward processing, we did not include measures of reward processing and reward discounting (14) for this analysis (note that we did include a measure of delay aversion in a subset of participants, but because it was not included in all youth it is excluded here). All of our measures are listed in Tables S4 and S5. Detailed explanation of each measure and how we obtained them are provided here.

Working memory and memory span. Three tasks were used to capture various components of working memory: (*i*) *Digit Span*: Youth completed the WISC-IV Digit Span forward (DSF) and backward (DSB) to assess verbal span and working memory abilities using standard procedures (4). Raw scores on the backward trials were retained for analyses. (*ii*) *Spatial Span*: Children next completed a computerized version of the Spatial Span subtest from the Wechsler Memory Scales (15) to examine visuospatial span and working memory capabilities. On this task, children were presented with a screen containing 10 squares arranged in a fixed position. Individual squares then changed color (from gray to yellow) in a fixed sequence. A tone sounded at the end of the sequence to note when the sequence was finished. Youth were then instructed to click on the squares in the order in which they changed color. The number of squares in the sequence began at three and increased to nine, with two trials for each sequence length. Similar to Digit Span, youth completed both Spatial Span forward (SSF) and backward (SSB) (i.e., recall the sequence in reverse order) versions of the task. (*iii*) *Stars Task*: We developed a computerized task modeled on work by Engel (16). For each trial of the task, presentations alternated between two types of trials: (*a*) "remember the number of stars," in which the screen displayed from one to three blue stars, and (*b*) "remember the location of the yellow star" in which the screen displayed five transparent stars, of which one was yellow. For the blue stars, participants were instructed to press the keyboard for the number of blue stars shown. For the yellow star presentations, children were instead instructed to remember its position within the row of five stars. Each block involved a series of alternating blue and yellow star presentations and ranged from one block of blue star/yellow star pairings (one-span set) to five blocks of blue star/yellow star pairings (five-span set). At the end of each set, youth were to circle on a corresponding page all of the positions in which they had viewed yellow stars, in the order in which they were presented. Children had to remember the position of one to five yellow stars. Five blocks of each span (one-, two-, three-, four-, and five-span) were presented for a total of 25 blocks. To score this task, one- and two-span trials were counted as indexing immediate recall whereas four- and five-span trials were considered as indexing working memory (blocks of 3 were omitted). To receive credit for a correct yellow star position, their response to the corresponding blue star trial also had to be correct to ensure they were engaged in the dual task design. Accuracy scores for each span length were then computed. The working memory trials (four- and five-span) were retained for analyses.

Interference control. DKEFS color word interference. This subtest from the DKEFS (17) (2001) was administered to assess interference control and is an analog to the classic Stroop task. Youth complete four conditions as part of this task. In the first condition (color naming: color word speed, CWSP) children were presented with a series of color patches on a page and instructed to name the colors out loud without skipping any or making any mistakes. In the second (word reading) they read aloud the color names as quickly as possible without making mistakes. In the third trial (inhibition/interference) youth viewed color names printed in different-colored ink and were to name the color of the ink (color word inhibition, CWIN). In the fourth (color word switching, CWSW), color names in contrasting ink colors appeared with or without a box around them. Youth were to name the color of the ink for those items with no box, but to read the word for those items in a box. The total completion times for each trial were retained for analyses.

Delis–Kaplan Executive Function System (DKEFS) trailmaking task. The DKEFS trailmaking task (17) (2001) was administered to assess cognitive-control and set-shifting abilities. Youth completed number and letter sequencing conditions (trails number and letter naming speed average, TRSP), and switching conditions (trails making task switching, TRSW). Number sequencing required youth to connect a series of numbers, in order (sequencing 1–16). Letter sequencing required connecting a series of letters, in alphabetical order (sequencing A–P). Switching required connecting numbers and letters in alternating sequence (1-A-2-B, etc.). The total completion times and total errors were recorded for each condition.

Response inhibition. The Stop Task (18) was administered to assess response inhibition and requires the suppression of a prepotent motor response. During this choice reaction time task, participants see an X or an O on a computer screen and respond rapidly with one of two keys to indicate which letter they had seen (called Go Response trials). In 25% of trials, a tone sounds shortly after the X or the O is displayed, indicating that participants are to withhold their response. A stochastic tracking procedure was used; stop signal reaction time (SSRT) was computed as an index of how much warning each participant needs to interrupt a response. Trials were presented across eight blocks of 32 trials. SSRT was calculated by subtracting the average stop signal delay from the average Go Response time (8, 18).

Response variability (SDX). The within-child variability of the reaction time in the Go Response trials was retained as a measure of response variability.

Arousal and activation. An identical pairs Continuous Performance test modified for children (19) was used to examine vigilance and sustained attention. In this task, children were presented with a rapid series of four-digit numbers. Youth were told to press a red button each time they saw a repeat of the exact digits (e.g., 2,524 followed by a second 2,524). The task was divided into five blocks of 288 paired trials. There were three pair types: (*i*) stim trials, or pairs of distinct digits (e.g., 6,923 and 2,524); (*ii*) catch trials, or digit pairs that differed by only one number (e.g., 2,524 and 2,534); and (*iii*) pair trials, or digit pairs that matched exactly (e.g., 2,524 and 2,524). Children received an accuracy score for each condition. To indicate arousal, we computed the signal detection index *d*-prime (20). The signal detection index *d*-prime (d') was computed for each of the five blocks. A higher *d*-prime score traditionally indicates greater sensitivity in distinguishing the targets (pair trials) from the nontargets (catch and stim trials). However, this score was reverse scored to ensure that higher scores for all measures indexed worse performance (weaker signal detection, presumed to be due to suboptimal arousal in the case of ADHD).

Temporal information processing. Tapping task. A computerized tapping task was administered to assess temporal information processing abilities, modified from that used by Toplak et al. (21). Youth were presented with either a visual or an auditory tapping rate and instructed to tap along at the same rate by pressing a red button. Two trials (slow rate of 1,000 ms between taps and fast rate of 400 ms between taps) for each presentation modality (visual and auditory) were administered to each child, for a total of four trials. The 1,000-ms condition is thought to have more memory demands than the 400-ms condition (22). A detrended SD was computed for each trial for each individual child to capture the extent to which each child's tapping rate varied against the target stimulus rate (400 ms or 1,000 ms) for each modality (visual and auditory). Larger values indicate greater deviation from the target tapping rate. These four detrended SDs were retained for analyses.

Response speed. As noted in Table S2, several variables measured aspects of response speed. We selected the subset of these measures that allowed the best fit for our confirmatory factor analysis (CFA) models—in this case, color naming and trail-making number sequencing. The other speed variables listed in Table S2 were omitted from the final analysis presented here.

Validity checks. All scores from each task were subjected to several validity criteria to ensure that the participants were completing the tasks correctly and that the data were providing an accurate measure of each neuropsychological construct. For example, in the continuous performance task (CPT) children's data had to show better accuracy in stim trials than in random trials for that block of data to be considered valid; in the stop task overall decision accuracy had to exceed 70% for that block of data to be considered valid. Data validity was coded as yes/no (1/0) for each task for each child. If a child failed to produce valid data on a given task, then his or her score for that task was estimated using full information maximum likelihood for the CFA and the computation of the factor scores for the modularity analysis. Similar quality checks were conducted on all of the data. In addition, we examined results of these checks to ensure that data excluded on the basis of these validity criteria did not differ between children on the basis of diagnostic group, sex, or age. All validity codes were unrelated to demographic (age, sex, ethnicity, grade) or diagnostic variables (ADHD and disruptive behavior disorder diagnosis, IQ; all $P > 0.15$).

Data Reduction for Neuropsychological Measures. A goal of our approach was to use a broad set of neuropsychological variables that would cover numerous domains that have been hypothesized to be relevant to ADHD. At the same time, we did not wish to use an excessive number of redundant indicators in our modularity analysis. We therefore sought to conduct rational reduction of the measures, according to the conceptual model that had guided our work as implied in Tables S4 and S5 of the main text. All measures were transformed such that higher scores were indicative of worse performance (e.g., slower speed or worse accuracy), so that all measures had the same valence. Fig. 2 of the main text portrays our primary conceptual model for how the variables were expected to relate. It also displays the factor loadings for the best-fitting seven-factor model. However, as noted in the main text *Methods*, because we were sensitive to the possibility of equivalent models, we also tested several competing models that conformed to our theorized reasons for choosing these measures. In the best-fitting six-factor model, we separated combined working memory and inhibition factors into a single "executive" factor. In the best-fitting five-factor model, we also combined "span" and "speed" factors into a single attention factor. In the best-fitting four-factor model, we also combined time reproduction with executive functioning.

Fit of all of these models was evaluated using several indexes, including the χ^2-value, the comparative fit index (CFI) (>0.90 = adequate), the Tucker Lewis index (TLI) (>0.90 = acceptable), and the root mean-square error of approximation (RMSEA) (<0.05 = good, 0.05–0.08 = adequate, 0.08–0.10 = marginal, >0.10 = poor). In these models, residual variances of measures from the same task were allowed to correlate (e.g., color naming, inhibition, and inhibition/switching from the DKEFS color word interference). Again, χ^2, CFI, TLI, and RMSEA were used to evaluate relative fit of the various models.

For the modularity analysis (below) factors were regressed for age and standardized across all participants to a mean of zero and SD of one, as we did not wish to cluster participants by age in this analysis and wanted all measures on the same metric scale.

Identification of Subgroups via Community Detection. To examine the strength of subject-to-subject relationships via graph theory, correlation matrices were created between subjects across the seven identified factor scores from the preceding feature reduction step. Each subject's factor scores were then correlated to every other subject's seven-factor scores. This procedure created two square correlation matrices (285 × 285 for ADHD and 213 × 213 for TDC) providing distance information (i.e., a correlation) between any given subject pair within the ADHD and TDC cohorts. Subsequent community detection was applied to these matrices separately.

Note that graph theoretic analyses, when applied to correlation matrices, generally rely on the thresholding of *r*-values. Typically, thresholding is a necessary step in the derivation of graphs (i.e., determining "connected" vs. "unconnected" pairs for either binary or weighted graphs). For our community detection procedures not every subject pair was deemed connected. The choice of threshold is therefore a critical decision point in the analytical process. For example, a choice of *r* approaching 1.0 will generate very sparse graphs, with a limited number of edges (i.e., few connected child pairs). In this instance "unattached" clusters of nodes could be deemed communities simply because of the sparse matrix. On the other hand a choice of *r* approaching 0.0 will generate densely connected graphs (i.e., nearly all child pairs would be deemed connected), where limited demarcations in the graph would be able to be identified. As such, to determine a proper threshold for which any two subjects were deemed connected or similar, we chose the maximum threshold where *reachability* remained equal to 1. Simply put, this is the maximum threshold where every subject is connected via at least one path to every other subject (no isolates). Thus, the graphs remain sparse, but fully connected (i.e., there are no isolated individuals lacking in any connections). This reachability threshold for the TDC graph was at $r = 0.56$, and the threshold for the ADHD graph was $r = 0.73$. However, to ensure our analysis did not depend on threshold selection, we also ran our community detection across multiple thresholds. In addition, we applied a weight-conserving modularity algorithm not dependent on thresholds (23). Both additional procedures yielded largely consistent results (Fig. S2).

Among the many methods used to detect communities in graphs, the modularity optimization algorithm of Newman is one of the most efficient (24). This method uses a quantitative measure of the observed vs. expected intracommunity connections, as a means to guide assignments of nodes (in this case subjects) into communities. We applied the modularity optimization algorithm to the group distance matrix noted above.

The strength of our modularity assignments was based on the quality index (Q), variation of information (VOI), and simulations created by repeating our analyses after 1,000 iterations of randomizing the factor scores across participants, thus generating a null distribution of Q (Fig. S3). Q of a graph is a quantitative measure of the number of edges found within communities vs. the number predicted in a random graph with equivalent degree distribution. A positive Q indicates that the number of intracommunity edges exceeds those predicted statistically. Q can range from −1.0 to +1.0, with 0 indicating there are no subgroups and 1.0 indicating perfectly reliable division of groups. A wide range of Q may be found for a graph, depending on how nodes are assigned to communities. The set of node assignments that returns the highest Q is the optimal community structure sought by the modularity optimization algorithm (for details see ref. 24).

The second index of robustness was VOI (25). As noted by Karrer et al. (25),

> Although it is true that networks with strong community structure have high modularity, it turns out that not all networks with high modularity have strong community structure. Indeed, there exist networks that most observers would consider to have no community structure at all that nonetheless have high modularity. . . . The reason for this at first peculiar finding is actually quite straightforward: the number of possible divisions of a network increases extremely fast with network size (faster than any exponential), so that although it is highly improbable that any one division will, purely by chance, have high modularity, it is, in the limit of large size, very likely that such a division will exist among the enormous number of possible candidates. As a result, high modularity is only a necessary but not sufficient condition for significant community structure.

As an example, Figs. S2*C* and S3*A* provide an instance from one of our randomized simulations that produced relatively high values of Q. Whereas the community pattern for the simulation does not replicate with the ADHD and TDC sample (as it did with the actual data), Q for this example was, nonetheless, high enough to suggest meaningful groupings. For this reason we also apply VOI as a secondary method for examining the robustness of our community assignments (25). For this analysis the modularity or community structure in the graph is compared using the same graph, but with a certain percentage (α) of random perturbation (or rewiring) of the edges/connections. Graphs with strong community assignments tend to remain the same even after perturbation, meaning there is little change in the VOI (in effect, the conclusions cannot be explained by any small portion of the data, suggesting results are not capitalizing on chance). This structure is distinct from a random network, wherein any community assignment is greatly affected by very little perturbation (i.e., results depend heavily on chance effects and are altered by any change in the data). By comparing the VOI for the actual data to the results for random data with the same parameters, we can evaluate whether our groupings do better than chance assignments. After computing the VOI across many different levels of α, one can quickly identify whether the community assignments of the experimental dataset deviate strongly from what might be expected in a random graph (Fig. S2 and ref. 25). We apply this method to both the true data and the randomized simulation. All of the preceding calculations were performed in MATLAB (Mathworks), using scripts generously provided by Olaf Sporns, Mikail Rubinov, and other collaborators (26).

Our last method to examine the robustness of our community structure was to use a more familiar approach. We generate a null distribution of Q on the basis of the weight-preserving modularity algorithm (thus without using thresholds) (23) via 1,000 randomizations of our participant factor scores (Fig. S2). We then use the Q-distribution to generate a z-score for measured modularity Q values of the form

$$z = \frac{Q - \mu}{\sigma},$$

which measures how many SDs or how significantly our Q values deviate from random. Although there is some criticism of this particular approach (25), it is generally robust and has support (27, 28), and we use it here only after having applied VOI, which controls for its deficiencies (25).

Support Vector Machine-Based Multivariate Pattern Analysis. Having formed our groups and tested their robustness, we proceeded to evaluate our ability to predict at the individual level. To do so, we used support vector machine (SVM)-based multivariate pattern analysis (MVPA) to identify whether (i) individual children can be classified into their group assignments and (ii) ADHD status could be discerned in the individual more effectively when these group assignments were considered.

SVM is a supervised classification algorithm rooted in statistical learning theory. Conceptually, input vectors are mapped to a higher-dimensional feature space using special nonlinear functions called kernels. Classification is performed by constructing a hyperplane in the feature space that optimally discriminates between two classes of the training data by maximizing the margin between two data clusters.

Given a training set of the form (x_i, y_i), where the vectors x_i are data points and y_i are the class labels, the SVMs require the solution to the optimization problem

$$\min_{w,b,\xi} \quad \frac{1}{2} w^T w + C \sum_{i=i}^{n} \xi_i,$$

subject to

$$y_i(w.x_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0,$$

where ξ_i are the slack variables, measuring the degree of a data point's misclassification, w are the weights defining the hyperplane, and $C > 0$ is the penalty parameter of the error term. The resultant decision function implemented by SVM can be written as

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{n} y_i \alpha_i K(x, x_i) + b\right),$$

where $K(x_i, x_j)$ is the kernel function. In our work, we use Radial basis kernel given by

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right).$$

SVMs are inherently two-class classifiers. Multiclass SVM aims to handle the K-class pattern classification problem by reducing the single multiclass problem into multiple binary classification problems. The most common method for such reduction is to build a set of *one-vs.-rest* binary classifiers that distinguish one of the classes from the rest. Another strategy is to build a set of *one-vs.-one* classifiers that distinguish between every pair of classes. For the *one-vs.-one* approach, classification is done by a max-wins voting strategy that chooses the class that is selected by the most classifiers. For the *one-vs.-rest* case (used in this work), classification of new instances is done by a winner-takes-all strategy, in which the classifier with the highest output function assigns the class. SVM classifications used a soft margin $C = 10$ and a radial basis function with $\sigma = 4$. We use Spider (http://people.kyb.tuebingen.mpg.de/spider/main.html), an object-orientated environment for machine learning in MATLAB, for generating the SVM models.

Subtype assignment. We first used SVM-based MVPA to determine, within a diagnostic group, how well individual children can be classified into their respective neuropsychological group assignments (first in TDC and subsequently in the ADHD cohort). For this determination we used a split replication procedure separately for both the TDC and the ADHD cohorts, dividing each sample on the basis of a balanced split procedure (Mahalanobis distance) into two groups (i.e., we divided the splits such that the two groups have the same N, age, sex, and IQ). Thus, for the ADHD group, there were 143 children in the first sample and 142 children in the second sample. For the TDC group, there were 107 in the first sample and 106 in the second sample.

We then applied our community detection procedure to each "split", which in a confirmatory fashion showed a similar break in the community structures within each split as it did in the whole samples (Fig. S4). From here we could then apply our SVM algorithm, using one split as our training set (i.e., to train our SVM) and one as our test set (i.e., to test the SVM's classification accuracy).

ADHD status. To test how well the ADHD status of our individual subjects could be determined using the neuropsychological measures alone within the community detection-determined groupings, we maximized the size of the training data by using a leave-one-out cross-validation (LOOCV) procedure. LOOCV involves removing a single subject as a test sample and then using the remaining data for feature selection and as the training set for the SVM predictor. This procedure is then repeated until each subject is used once as the test case. LOOCV is a commonly implemented cross-validation tool because it maximizes the amount of data used for training, is widely used in machine learning (29), and has been shown to provide a conservative estimate of a classifier's or predictor's true accuracy. We first used this procedure to identify how well an individual subject could be predicted when considering the ADHD and TDC populations as homogeneous groups. We then repeated the procedure, looking within community types (i.e., subgroups 1–4) (*Results* and Figs. 3 and 4 in main text) to determine whether the inferential power of ADHD classification is improved by examining within each profile type.

1. Martel M, Nikolas M, Nigg JT (2007) Executive function in adolescents with ADHD. *J Am Acad Child Adolesc Psychiatry* 46:1437–1444.
2. Puig-Antich J, Ryan N (1986) *The Schedule for Affective Disorders and Schizophrenia for School-Age Children* (Western Psychiatric Institute and Clinic, Pittsburgh).
3. Conners CK (2008) *Conners Manual* (Multi-Health Systems, Toronto), 3rd Ed.
4. Wechsler D (2003) *Wechsler Intelligence Scale for Children Technical and Interpretive Manual* (The Psychological Corporation, San Antonio), 4th Ed.
5. Wechsler D (2005) *Wechsler Individual Achievement Test* (Harcourt Assessment, London).
6. Martinussen R, Hayden J, Hogg-Johnson S, Tannock R (2005) A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 44:377–384.
7. Barkley RA (1997) Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychol Bull* 121:65–94.
8. Nigg JT (1999) The ADHD response-inhibition deficit as measured by the stop task: Replication with DSM-IV combined type, extension, and qualification. *J Abnorm Child Psychol* 27:393–402.
9. Castellanos FX, Tannock R (2002) Neuroscience of attention-deficit/hyperactivity disorder: The search for endophenotypes. *Nat Rev Neurosci* 3:617–628.
10. Toplak ME, Dockstader C, Tannock R (2006) Temporal information processing in ADHD: Findings to date and new methods. *J Neurosci Methods* 151:15–29.
11. Sergeant J (2000) The cognitive-energetic model: An empirical approach to attention-deficit hyperactivity disorder. *Neurosci Biobehav Rev* 24:7–12.
12. Welsh MC, Pennington BF (1988) Assessing frontal lobe functioning in children: Views from developmental psychology. *Dev Neuropsychol* 4:199–230.
13. Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF (2005) Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biol Psychiatry* 57:1336–1346.
14. Sagvolden T, Johansen EB, Aase H, Russell VA (2005) A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behav Brain Sci* 28(3):397–419; discussion 419–468.
15. Wechsler D (1987) *Wechsler Memory Scale-Revised Manual* (The Psychological Corporation, San Antonio).
16. Conway AR, et al. (2005) Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev* 12:769–786.
17. Delis DC, Kaplan E, Kramer J (2001) *Delis Kaplan Executive Function System* (The Psychological Corporation, San Antonio).
18. Logan GD, Schachar RJ, Tannock R (1997) Impulsivity and inhibition control. *Psychol Sci* 8:60–64.
19. Curko Kera EA, Marks DJ, Berwid OG, Santra A, Halperin JM (2004) Self-report and objective measures of ADHD-related behaviors in parents of preschool children at risk for ADHD. *CNS Spectr* 9:639–647.
20. Sergeant JA, Oosterlaan J, van der Meere JJ (1999) *Information Processing and Energetic Factors in Attention Deficit/Hyperactivity Disorder* (Kluwer/Plenum, New York).
21. Toplak ME, Tannock R (2005) Tapping and anticipation performance in attention deficit hyperactivity disorder. *Percept Mot Skills* 100:659–675.
22. Ivry RB (1996) The representation of temporal information in perception and motor control. *Curr Opin Neurobiol* 6:851–857.

23. Rubinov M, Sporns O (2011) Weight-conserving characterization of complex functional brain networks. *Neuroimage* 56:2068–2079.
24. Newman ME (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103:8577–8582.
25. Karrer B, Levina E, Newman ME (2008) Robustness of community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 77:046119-1–046119-9.
26. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 52:1059–1069.
27. Guimerà R, Sales-Pardo M, Amaral LA (2004) Modularity from fluctuations in random graphs and complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 70:025101-1–025101-8.
28. Reichardt J, Bornholdt S (2007) Partitioning and modularity of graphs with arbitrary degree distribution. *Phys Rev E Stat Nonlin Soft Matter Phys* 76:015102-1–015102-4.
29. Christianini N, Shawe-Taylor J (2006) *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge Univ Press, Cambridge, UK), 10th Ed.

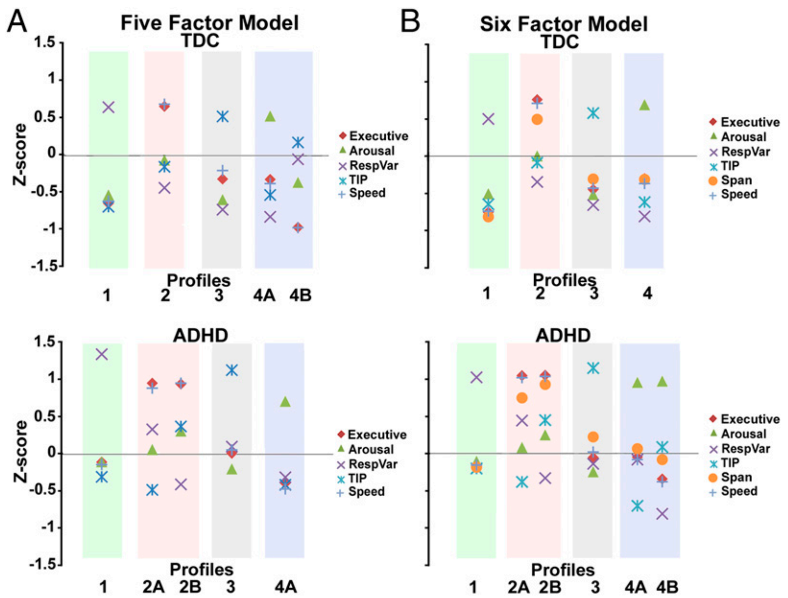

Fig. S1. Community detection applied to alternative five- and six-factor models. As noted in the main text because a limitation of confirmatory factor analysis is the potential for untested models to have equivalent fit, we tested several competing models for our feature reduction step. (*A*) Community detection based on the five-factor model. (*B*) Community detection based on the six-factor model. The main findings in the text did not deviate widely on the basis of the three best-fitting models.

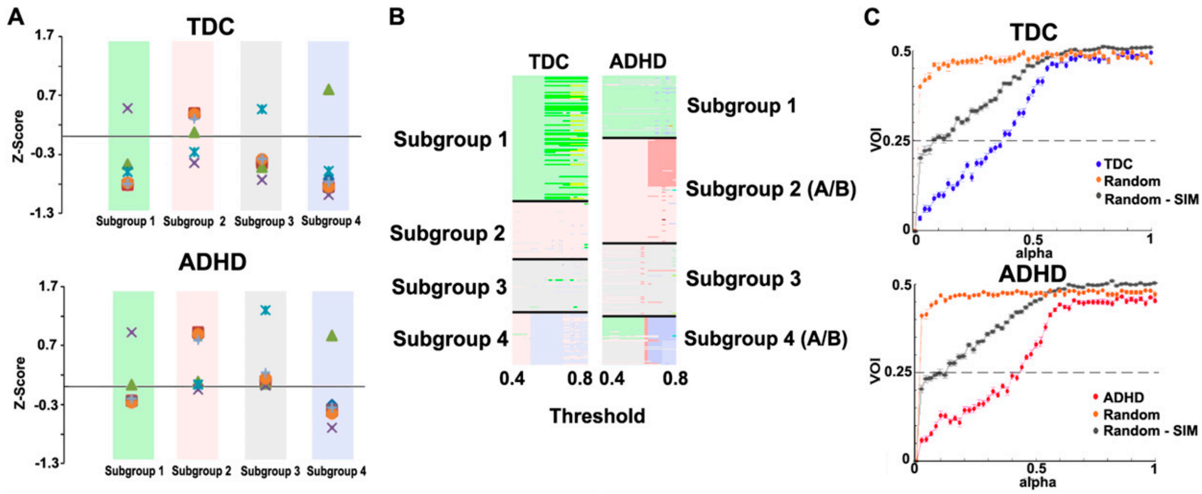

Fig. S2. (*A*) To ensure our subgroups in both the ADHD and the TDC populations were not specific to a particular threshold, we reexamined our data using a weight-preserving modularity algorithm (23). Our subgroupings in both the TDC and the ADHD populations replicated using this procedure. *Q* values were high (*Q* = 0.5 TDC and 0.51 ADHD). The one deviation was in regard to the ADHD population, where profiles 2A and 2B as well as profiles 4A and 4B merged into one profile as in the TDC population. (*B*) We also reexamined our subgroupings across multiple thresholds, as a secondary measure to show the communities identified were largely independent of threshold selection. Here thresholds were applied from *r* = 0.4 to 0.8. The color codes correspond to profile colors as in *A*. The *y* axis represents participants. Although there are small deviations across thresholds, group assignments remain largely similar across thresholds. Not surprisingly, for the TDC population as the thresholds increase past 100% reachability (i.e., *r* = 0.53), the variability in community assignments increases due to nodes and clusters of nodes becoming "disengaged" in the network. (*C*) Here we examine variation of information (VOI) across the identified communities as a way to reveal a structure highly deviant from random. VOI is a measure of how much information is not shared between two sets of community assignments and allows for the quantification of network robustness (1, 2). Values of 0 indicate identical community assignments, and values of 1 indicate maximally different community assignments. To assess the stability of community assignments, the edges of a network are randomized with probability α to perturb the network, and the VOI between the original and perturbed networks is calculated over a range of α. A random network with equivalent degree and a network with high *Q* based on our random simulations (Fig. S3*A*) were generated for comparison. The entire perturbation process was repeated 20 times for each network. Compared with the random graphs, the community assignments in both ADHD and TDC are significantly robust.

1. Karrer B, Levina E, Newman ME (2008) Robustness of community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 77:046119-1–046119-9.
2. Meila M (2007) Comparing clusterings - an information based distance. *J Multivariate Anal* 98:873–895.

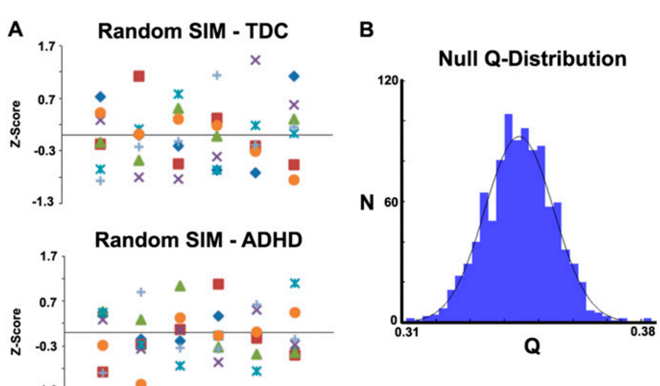

Fig. S3. Generating a null distribution of *Q* on the basis of the weight-preserving modularity algorithm introduced by Rubinov and Sporns (23). (*A*) One example of our simulations created by repeating our analyses after 1,000 iterations of randomizing the factor scores across participants. Whereas using a threshold of 100% reachability yields an average of 11 communities in each cohort for these randomizations, for visualization and explanatory purposes we reduced the threshold for this example, such that only 6 communities were present. Perhaps not surprisingly, the cohort community patterns do not replicate across cohorts and factor distributions are highly varied. Even so, for this example from our 1,000 random iterations *Q* values were high (*Q* = 0.55 ADHD and *Q* = 0.58 TDC) after thresholding, and moderately high using the weight-preserving modularity algorithm by Rubinov and Sporns (23) (*Q* = 0.36 ADHD and *Q* = 0.38 TDC). As noted in *SI Text*, this result might occur in some random graphs and thus highlights the importance of secondary means by which to test the robustness of modularity assignments (e.g., VOI, within-sample replication, *z*-score relative to null distribution, as used here). (*B*) Full *Q* null distribution based on the 1,000-factor randomizations, using the weight-preserving algorithm used in Fig. S2*A*. As this modularity method does not rely on thresholding of *r*-values, it is more easily calibrated for our networks to generate a null distribution. This distribution was used to generate *z*-scores for the *Q* values in Fig. S2*A* on the basis of our original matrices (actual data). For the TDC cohort *Z* = 15.51. For the ADHD cohort *Z* = 16.51.

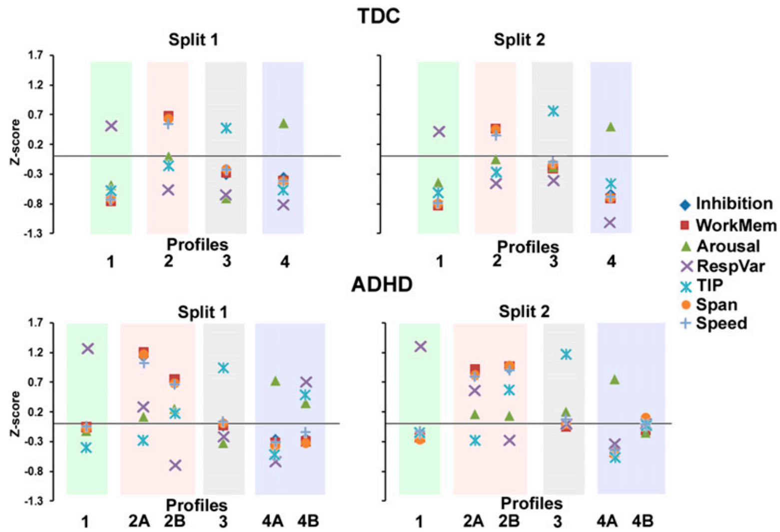

Fig. S4. A split replication procedure was applied for the SVM-based MVPA analysis. To use our SVM to test the ability for single-subject classification into any given community we split the data into a test and a training set. Community detection was applied to each split separately and revealed similar findings. The one deviation for the full dataset was subgroup 4B in the ADHD set. The overall pattern appeared deviant when compared using the full ADHD sample, and only one subject localized to this group in the second split. Because of this limitation, classification for this particular group was not applied.

Table S1. Descriptive and demographic statistics

	Control	ADHD-PI	ADHD-C	*P*	Pairwise differences
N	213	107	137		
% Male	42.3	57	75.9	<.001	Control<PI<C
% Caucasian	76.5	74.8	70.8	0.49	—
% African-American	8.9	5.6	9.5	0.5	—
% Latino	3.8	2.8	8.8	0.05	—
% Mixed/biracial	9.4	13.1	10.9	0.6	—
Age (SD)	11.0 (2.4)	11.3 (2.3)	9.9 (2.2)	<.001	C<PI=Control
Income (SD)*	79.0 (47.4)	75.2 (36.4)	55.9 (38.0)	<.001	C<PI=Control
Full-scale IQ (SD)	107.12 (14.14)	102.79 (16.18)	102.69 (14.33)	0.01	C=PI<Control
% Stimulant medication	1.4	28	45.3	<.001	Control<PI<C
KSAD diagnostics					
Inattention symptoms (SD)	.74 (1.4)	7.4 (1.5)	7.6 (1.6)	<.001	Control<PI=C
Hyperactive symptoms (SD)	.62 (1.2)	1.8 (1.7)	6.2 (2.0)	<.001	Control<PI<C
% ODD	5.2	18.9	38.7	<.001	Control<PI<C
% CD	0	2.8	5.8	0.01	Control <C
Conners parent oppositionality	48.3 (9.5)	55.9 (12.7)	63.8 (13.2)	<.001	Control<PI<C
Conners parent cognitive problems	48.2 (7.6)	72.8 (10.2)	69.9 (10.7)	<.001	Control<C<PI
Conners parent hyperactivity	48.9 (8.2)	57.2 (11.9)	71.2 (12.3)	<.001	Control<PI<C
Conners teacher oppostionality	48.5 (7.3)	52.0 (10.1)	58.4 (13.6)	<.001	Control<PI<C
Conners teacher cognitive problems	49.6 (8.5)	66.7 (11.0)	68.5 (13.4)	<.001	Control<PI=C
Conners teacher hyperactivity	49.6 (8.8)	57.7 (10.5)	62.2 (12.2)	<.001	Control<PI=C

P values indicate 3-group significance test. Conners' *T* scores and standard deviations provided reflect age and sex norms.
*Income reported in thousands.

Table S2. Neuropsychological measures, variables, and constructs

Task	Variables	Construct
Spatial span	Backward total correct	Working memory
	Forward total correct	Encoding, span
Digit span	Backward total correct	Working memory
	Forward total correct	Encoding, span
Star span	Long trials total correct	Working memory
	Short trials total correct	Encoding, span
DKEF color word	Color naming time	Naming speed
	Word reading time	Reading speed*
	Inhibition/interference time	Interference control, executive function
	Inhibition switch	Switching speed, executive function
DKEF trailmaking	Visual scanning time	Scan speed*
	Number sequencing time	Speed
	Letter sequencing time	Speed
	Number letter time	Switching speed, executive function
Stop task	Go reaction time	Decision speed/speed
	Reaction time SD	Response variability
	Stop signal reaction time	Response suppression/inhibition
CPT	*d*-prime	Arousal
Tapping task	400-ms detrended sd visual	Time reproduction
	400-ms detrended sd auditory	Time reproduction
	1,000-ms detrended sd visual	Time reproduction
	1,000-ms detrended sd visual	Time reproduction

CPT, continuous performance task; DKEF, Delis–Kaplan executive function battery; sd, standard deviation.
*Variables were excluded from final confirmatory factor model to improve model fit.

Table S3. Subgroup comparisons between ADHD and TDC participants

All TDC vs. All ADHD	ANOVA: source	*F*	Prob > *F*	Post hoc *t* test	Inhibition	WorkMem	Arousal	RespVar	TIP	Span	Speed
Comparison: TDC vs. ADHD	Diagnosis	250.52	0.000	*P*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Behavior	0.05	0.999	*t*-stat	7.324	7.391	3.954	3.689	5.388	6.907	7.403
	Diagnosis*behavior	2.43	0.024	SD	0.951	0.950	0.986	0.988	0.973	0.956	0.950
Within profiles											
Subgroup comparison: profile 1	Diagnosis	106.84	0.000	*P*	0.000	0.000	0.003	0.000	0.012	0.000	0.000
	Behavior	46.81	0.000	*t*-stat	4.363	4.452	3.070	4.462	2.535	4.204	4.440
	Diagnosis*behavior	0.64	0.697	SD	0.798	0.767	0.915	0.928	0.810	0.741	0.809
Subgroup comparison: profile 2A	Diagnosis	16.74	0.000	*P*	0.128	0.167	0.005	0.257	0.001	0.190	0.104
	Behavior	29.01	0.000	*t*-stat	1.535	1.395	2.885	−1.141	3.288	1.321	1.642
	Diagnosis*behavior	1.80	0.097	SD	0.822	0.807	0.719	0.759	0.765	0.846	0.836
Subgroup comparison: profile 2B	Diagnosis	25.59	0.000	*P*	0.033	0.022	0.276	0.000	0.492	0.050	0.039
	Behavior	26.88	0.000	*t*-stat	2.165	2.333	1.096	4.522	−0.689	1.982	2.095
	Diagnosis*behavior	1.96	0.069	SD	0.929	0.914	0.737	0.733	0.806	1.006	0.924
Subgroup comparison: profile 3	Diagnosis	55.92	0.000	*P*	0.007	0.007	0.173	0.000	0.002	0.011	0.011
	Behavior	24.56	0.000	*t*-stat	2.750	2.750	1.371	4.955	3.225	2.592	2.596
	Diagnosis*behavior	1.10	0.359	SD	0.748	0.721	1.017	0.804	0.967	0.724	0.835
Subgroup comparison: profile 4A	Diagnosis	13.59	0.000	*P*	0.124	0.087	0.512	0.000	0.490	0.190	0.130
	Behavior	18.87	0.000	*t*-stat	1.558	1.742	0.660	4.207	−0.695	1.326	1.535
	Diagnosis*behavior	1.60	0.146	SD	0.757	0.720	1.125	0.693	0.686	0.685	0.771
Subgroup comparison: profile 4B	Diagnosis	0.99	0.321	*P*	0.844	0.799	0.796	0.754	0.005	0.687	0.902
	Behavior	19.47	0.000	*t*-stat	−0.197	−0.255	0.259	0.314	2.882	−0.405	0.123
	Diagnosis*behavior	1.07	0.377	SD	0.778	0.752	1.159	0.701	0.753	0.713	0.795

Table S4. TDC cognitive profile (or subgroup) demographics

TDC	Profile 1	Profile 2	Profile 3	Profile 4	*P* value
N	92	43	39	39	—
% male	45.65	46.51	38.46	33.33	—
% female	54.35	53.49	61.54	66.67	—
Age	11.38	11.50	11.74	11.57	0.89
FSIQ	109.64	100.88*	106.03	109.15	0.01
INT	0.79	0.70	0.63	0.74	0.94
HYP	0.54	0.62	0.79	0.65	0.72
% of sample	43.19	20.19	18.31	18.31	—

P values indicate *N*-group significance test. Profile 2 for both ADHD and TDC had small but significant lower IQs relative to other profiles. FSIQ, full scale IQ; INT, average number of inattentive symptoms; HYP, average number of hyperactive symptoms.

*Post hoc significant pairwise differences at $P < 0.05$ (profiles 1 and 4).

Table S5. ADHD cognitive profile (or subgroup) demographics

ADHD	Profile 1	Profile 2A	Profile 2B	Profile 3	Profile 4A	Profile 4B	*P* value
N	61	49	56	71	24	24	—
% male	54.10	77.55	80.36	56.34	54.17	62.50	—
% female	45.90	22.45	19.64	43.66	45.83	37.50	—
Age	11.34	11.54	10.33	11.01	11.69	10.84	0.06
FSIQ	104.21	96.45*	99.07†	106.32	102.00	107.96	0.00
INT	6.32	7.23	7.00	6.84	6.65	7.06	0.80
HYP	3.84	3.68	3.91	4.08	3.98	5.35	0.31
% of sample	21.40	17.19	19.65	24.91	8.42	8.42	—

P values indicate *N*-group significance test. Profile 2 for both ADHD and TDC had small but significant lower IQs relative to other profiles.

*Post hoc significant pairwise differences at $P < 0.05$ (profiles 3 and 4A).

†Post hoc significant pairwise differences at $P < 0.05$ (profiles 1, 3, and 4A).