Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD

Damien A. Fair[a,b,c,1], Deepti Bathula[a,d], Molly A. Nikolas[e], and Joel T. Nigg[a,b]

Departments of [a]Behavioral Neuroscience and [b]Psychiatry and [c]Advanced Imaging Research Center, Oregon Health and Science University, Portland, OR 97239; [d]Department of Computer Science and Engineering, Indian Institute of Technology Ropar, Rupagnar, India 140001; and [e]Department of Psychology, University of Iowa, Iowa City, IA 52242

Edited by James Swanson, University of California, Irvine, CA, and accepted by the Editorial Board February 24, 2012 (received for review September 23, 2011)

Research and clinical investigations in psychiatry largely rely on the de facto assumption that the diagnostic categories identified in the *Diagnostic and Statistical Manual* (DSM) represent homogeneous syndromes. However, the mechanistic heterogeneity that potentially underlies the existing classification scheme might limit discovery of etiology for most developmental psychiatric disorders. Another, perhaps less palpable, reality may also be interfering with progress—heterogeneity in typically developing populations. In this report we attempt to clarify neuropsychological heterogeneity in a large dataset of typically developing youth and youth with attention deficit/hyperactivity disorder (ADHD), using graph theory and community detection. We sought to determine whether data-driven neuropsychological subtypes could be discerned in children with and without the disorder. Because individual classification is the sine qua non for eventual clinical translation, we also apply support vector machine-based multivariate pattern analysis to identify how well ADHD status in individual children can be identified as defined by the community detection delineated subtypes. The analysis yielded several unique, but similar subtypes across *both* populations. Just as importantly, comparing typically developing children with ADHD children within each of these distinct subgroups increased diagnostic accuracy. Two important principles were identified that have the potential to advance our understanding of typical development and developmental neuropsychiatric disorders. The first tenet suggests that typically developing children can be classified into distinct neuropsychological subgroups with high precision. The second tenet proposes that some of the heterogeneity in individuals with ADHD might be "nested" in this normal variation.

psychiatric disorders | research domain criteria | cognition | modularity | executive functions

In psychiatry, research and clinical investigation largely relies on the de facto assumption that the diagnostic categories identified in the *Diagnostic and Statistical Manual* of mental disorders (DSM-IV) represent etiologically homogeneous syndromes. However, there is considerable evidence that suggests the DSM does not necessarily describe homogenous conditions, but rather reflects the end result of multiple unique independent mechanistic pathways within a given disorder (1, 2). The mechanistic heterogeneity that potentially underlies the existing classification scheme might be limiting our ability to clarify etiology and identify novel therapeutics for several psychiatric illnesses (3).

A salient example, and our focus here, is attention deficit/hyperactivity disorder (ADHD). It is one of the earliest onset, most common, and costly neurodevelopmental disorders in child psychiatry (4, 5). Until recently, causal models of ADHD, as with other mental disorders, proposed a single core dysfunction (6). Investigators typically compare a group of children with ADHD defined by core symptoms (i.e., DSM) to a group of control children without the disorder. Statistical group differences based on psychometrics, functional brain imaging, or genetics are then used to inform models of the disorder.

This assumption of homogeneity in the case of ADHD has been questioned in numerous theoretical papers (7–13). For example, Nigg et al. (13) showed that several neuropsychological measures central to ADHD had substantial distributional overlap between ADHD and control samples. The data suggested that only a small minority of subjects with the disorder could be considered clinically "affected" on the basis of any one measure (13). Similar findings have been noted elsewhere (14, 15). In other words, whereas numerous unique neuropsychological measures have been proposed as related to ADHD, perhaps each of them applies to only a subset of those subjects with the disorder.

Although the role of heterogeneity in clinical populations has caught the collective attention of funding agencies (16) and the scientific community (8, 13, 15, 17), another, perhaps less palpable, reality may also be interfering with progress in understanding psychiatric illnesses—heterogeneity in typical populations. In the same way that investigators are often bound by the "cognitive box" (1) of the DSM when examining atypical development, they are also generally obliged to conduct their analysis as if typically developing comparison populations represent a monolithic group. Although consensus remains elusive on defining specific personality or cognitive "types" (18, 19), substantial evidence has accrued that individual differences in successful adaptive psychological styles are central to human development, functioning, social cohesion, and health outcomes (20–23). It may be that identifying a mechanism associated with a mental disorder requires comparing individuals to well-adjusted persons with the same cognitive style or profile.

Although it is easy to propose conceptually that there must be distinct subgroups within mental disorders (or typical populations), empirically demonstrating such subgroups is not straightforward. In the case of ADHD, emphasis has been on latent class analysis using symptom profiles (24), personality traits (25), and developmental trajectories of symptoms (26). These approaches are promising but appear to have mainly tended to identify severity classes rather than distinct categories (27). Efforts to identify types using neurocognitive measures—in theory, related to pathophysiological mechanisms—are still in the beginning stages. A key goal of this work is to identify procedures that are not prone to simply identifying severity groupings.

One approach, which may prove fruitful toward this goal, emanates from graph theory. Graph theory is a mathematical discipline about the study of networks, in which networks are

Author contributions: D.A.F., D.B., M.A.N., and J.T.N. designed research; D.A.F., D.B., M.A.N., and J.T.N. performed research; D.A.F., D.B., M.A.N., and J.T.N. contributed new reagents/analytic tools; D.A.F., D.B., M.A.N., and J.T.N. analyzed data; and D.A.F. and J.T.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.S. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: damien.fair@aya.yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1115365109/-/DCSupplemental.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

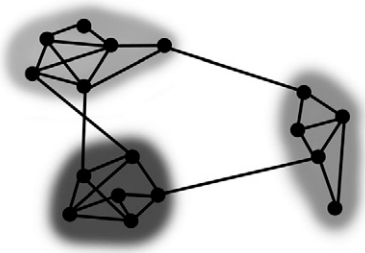

Fig. 1. Graph theory and community detection. Displayed is a depiction of a network, where nodes (solid circles) are connected by edges (solid lines). In this paper nodes are participants and edges are correlations between participants' neuropsychological scores. Community detection algorithms (29) can be applied to graph structures to identify clusters of nodes (shaded clouds) that share many edges within clusters relative to between clusters.

simply sets of nodes or vertices joined in pairs by lines or edges (Fig. 1). Graph theory has been used to examine the organization of a number of relationships within systems (28). Importantly, many systems have well-defined internal structures and can be described or demarcated with graph theoretical analyses (28, 29). One area that has received considerable attention is the detection of community structure in networks. Community structure refers to the appearance of densely connected groups of nodes, with only sparse connections between the groups (Fig. 1) (29). The focus of this report is whether groups of children that share similar empirical neuropsychological features segregate to form specific data-driven phenotypic subtypes.

Because individual classification is the *sine qua non* for eventual translation to clinical use, we followed our community detection analysis with an investigation using support vector machine (SVM)-based multivariate pattern analysis (MVPA) (30, 31) to identify how well individual children can be identified as defined by the community detection delineated subtypes. SVMs are supervised classification algorithms rooted in statistical learning theory, capable of recognizing patterns for the purposes of categorization. Typically SVMs examine a set of training data for which each data point (e.g., person) has been assigned to a unique category with several defining features. On the basis of patterns among the features within each category, the training algorithm then builds a model capable of assigning new data points (e.g., individuals) into these specific categories. Here we use SVM-based MVPA to determine whether there is sufficient information in the neuropsychological scores to predict whether any individual can be accurately classified into a particular neuropsychological subgroup or profile defined by the community detection procedure. We also use the approach to determine whether ADHD status can be more accurately assigned after considering the community-based profiles.

Results

Feature Reduction Supported a Seven-Factor Model of Cognitive Abilities. For the current investigation we apply community detection to a well-characterized dataset of 498 children who include both typically developing control youth (TDC) (n = 213) and youth with ADHD (n = 285) (Table S1). From these youth, some 20 neuropsychological measures were obtained that were intended to cover a wide domain of cognitive functions variously theorized to be involved in ADHD (Table S2 and Fig. 2).

Our approach was to use a broad set of neuropsychological variables relevant to ADHD, while avoiding use of an excessive number of redundant indicators in our analysis. We therefore conducted a rational feature reduction using confirmatory factor analysis (CFA), according to the conceptual model that had guided our work (Fig. 2 and refs. 32 and 33). All measures were transformed such that higher scores were indicative of worse

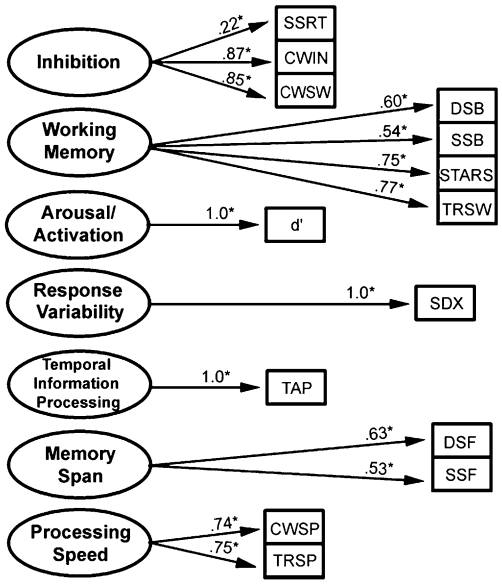

Fig. 2. Data reduction for neuropsychological measures. Confirmatory factor analysis (CFA) was used to conduct rational reduction of the measures listed in Table S2. Shown is our conceptual model that depicts how we hypothesized that our measured variables relate to seven latent factors. It also displays the factor loadings for the seven-factor model. For ease of presentation, the figure does not display error terms, cross loadings, or correlations among latent factors. CWIN, color word inhibition; CWSP, color word speed; CWSW, color word switching; d', D-prime; DSB, digit span backward; DSF, digit span forward; SDX, response variability; SSB, spatial span backward; SSF, spatial span forward; SSRT, stop signal reaction time; STARS, stars task; TAP, tapping task (temporal information processing task); TRSP, trails number and letter naming speed average; TRSW, trails-making task switching.

performance (e.g., slower speed or worse accuracy). Fig. 2 portrays our primary model with the empirical factor loadings (*SI Text*). Because we were sensitive to the possibility of equivalent models, we also tested several competing models that conformed in varying degrees to our theorized reasons for choosing these measures. Fit of all models was evaluated using several indexes, including the χ^2-value, the comparative fit index (CFI), the Tucker Lewis index (TLI), and the root mean-square error of approximation (RMSEA) (*SI Text*).

Results for the feature reduction for the neuropsychological measures showed that a one-factor model fit inadequately [χ^2 (70) = 386.2, CFI = 0.87, TLI = 0.84, RMSEA = 0.09], justifying our effort to create a multiconstruct model. Fit was satisfactory and comparable for the best-fitting five-, six-, and seven-factor models as follows: five-factor χ^2 (63) = 103.4, CFI = 0.98, TLI = 0.97, RMSEA = 0.036; six-factor χ^2 (58) = 95.1, CFI = 0.98, TLI = 0.98, RMSEA = 0.035; and seven-factor χ^2 (52) = 89.9, CFI = 0.98, TLI = 0.97, RMSEA = 0.038. For our main analysis we present results for the seven-factor model because we viewed it as conceptually the most differentiated description of the abilities tested, the most consistent with our conceptual framework, and equivalent in fit to the other models. However, we conducted the modularity analysis using the best-fitting six- and five-factor solutions as well to see whether community assignments were conceptually similar with these slightly different indicators, and indeed they were (Fig. S1).

Comparing the Full ADHD and TDC Cohorts Replicated Previous Findings. The second stage of our analysis began with a traditional comparison between the ADHD and TDC cohorts. This analysis was followed by our SVM pattern classifier to determine how well the neuropsychological scores can inform individual distinctions between ADHD and TDC.

The comparisons yielded significant differences across all measures (Fig. 3*A* and Table S3). Importantly, despite these highly reliable differences, the SVM classifier was unable to make strong distinctions in individual cases with regard to ADHD status (65% accuracy, 86.3% sensitivity, and 38.5% specificity).

This analysis served as a baseline to determine whether the subgroups identified by our community detection analysis (below) showed similar findings or whether the atypical nature of any given factor resided in specific profiles (i.e., groups of participants). In the same way, we also used this baseline to determine whether diagnostic classification could be improved when considering our community detection-derived profiles.

Community Detection Revealed Unique Subpopulations or Profiles in both the TDC and the ADHD Cohorts. We first applied community detection to the TDC population. Although common practice would expect this investigation to yield one unitary group, the analysis instead yielded four unique communities or subgroups (Fig. 4). Importantly, the quality index ($Q = 0.45$) and variation of information (VOI) (two measures of community robustness), as well as a secondary randomization analysis (*SI Text* and Figs. S2 and S3), showed that the subgroups identified here were significantly different from random.

Each of the four groups had unique patterns of factor scores. One group (43% of sample) appeared to have a pattern consistent with more response variability relative to their peers (subgroup 1). The second group (20% of sample) had reduced working memory, memory span, inhibition, and output speed (subgroup 2). The third group (18% of sample) had relatively inaccurate temporal information processing (subgroup 3). The last group (18% of sample) had relatively weak signal detection, suggesting suboptimal or altered arousal (subgroup 4). These groups show minimal differences in intelligence quotient (IQ), age, or sex ratios (with one exception, see Tables S4 and S5). This analysis shows that despite traditional assumptions that nonclinical control populations are uniform entities, our results fit better with the alternative supposition that there are unique neuropsychological profiles even in typically developing, well-adjusted samples.

We next applied the same community detection procedure to the ADHD sample. Independently testing the community structure in this sample revealed similar findings, albeit via six groups (Fig. 4). Again, Q ($Q = 0.55$) and the VOI analysis (Fig. S2) highlighted the robustness of the communities. Similar to the TDC, the first group (21% of sample) appeared to have a pattern consistent with high levels of reaction time variability relative to their peers (subgroup 1). As with TDC, the second group (17% of sample) appeared to have reduced working memory, memory span, and output speed (i.e., subgroup 2A). The third group (20% of sample) also had the same apparent weaknesses, but with a slightly modified profile in the remaining measures (i.e., subgroup 2B). The fourth group (25% of sample) had inaccurate temporal information processing (subgroup 3). The fifth group (8% of sample) was also similar to the fourth group of TDC youth in that they appeared to be characterized by suboptimal arousal (subgroup 4A). The sixth group (8% of sample) was similar to the fifth with regard to low arousal, however, with a slightly different profile in the remaining measures (subgroup 4B).

Strikingly, even though these groups were all distinct in their neuropsychological profiles, they again showed minimal differences in symptom scores, IQ, age, or sex ratios (Tables S4 and S5). Thus, these are not simply more or less severe ADHD groups, but rather unique cognitive profiles within children who all have similar severity of ADHD.

SVM-Based MVPA Highlights the Robustness of the Community Detection-Defined Profiles. To further test the overall robustness of these defined groups we next arranged our SVM to classify individual subjects between the profiles identified within the TDC and ADHD cohorts. To do this we first split the ADHD and TDC samples into two: each cohort having a test set and a training set (*SI Text*). We independently reapplied our community detection procedure to both the ADHD and the TDC cohorts and reproduced the group assignments identified in the first analysis (Fig. S4). We then used our training set to train the SVM on the four group assignments for TDC and the six group assignments for the ADHD populations. The test evaluates how robustly any one individual can be classified into his/her neuropsychological profile.

Classification generally was quite respectable across the groups with 78% accuracy for the TDC population (subgroup 1, 93%; subgroup 2, 71%; subgroup 3, 71%; subgroup 4, 81.25%) and 77% accuracy for the ADHD population (subgroup 1, 78%; subgroup 2A, 75%; subgroup 2B, 66.25%; subgroup 3, 77%; subgroup 4A, 88%; note that subgroup 4B was unable to be examined as it was not reproduced in large enough numbers of participants within each split to do SVM testing) (*SI Text*).

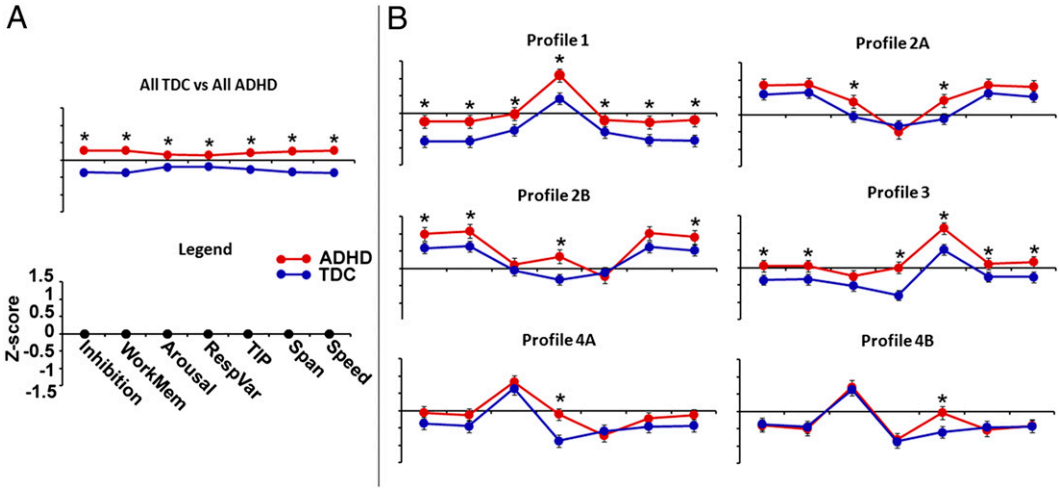

Fig. 3. Atypical neuropsychological measures are specific to cognitive subgroup. Here we show the comparison of neuropsychological measures between ADHD and TDC. (*A*) Comparison between the entire TDC and ADHD samples. (*B*) ADHD vs. TDC comparison within each subgroup. Interestingly, atypical neuropsychological measures relative to the control population are not uniform across all subgroups. Rather, each subgroup has a unique pattern of atypical measures (*, significant differences between groups; details in Table S3).

Atypical Neuropsychological Measures Are Unique on the Basis of Profile and ADHD Classification Improves Within Profiles. Considering the unique nature of the subclassifications, we next returned to our group comparisons between ADHD and TDC. However, instead of comparing all ADHD with all TDC, we now compared ADHD and TDC neuropsychological scores within their specific assigned neuropsychological "type" or group. The results are presented in Fig. 3 and Table S3. Whereas the ANOVA in six of the seven factors showed a main effect of diagnosis, the post hoc comparisons showed that not all subgroups had the same atypical neuropsychological scores.

Among ADHD subgroups, atypical inhibitory control, speed, and working memory were each found in three groups. Atypical arousal and span were each identified in two groups. Atypical response variability and temporal information priming were each found in four groups. Interestingly, the ADHD classification improved or remained the same for all but one of the groups, relative to the initial All ADHD vs. All TDC analysis. Group 4A had the maximal improvement at 84.1% classification of ADHD vs. TCD youth (83.3% sensitivity and 84.6% specificity). The remaining groups were as follows: group 1, 68% total, 49.2% sensitivity, 80.4% specificity; group 2A, 68.5%, 83.7%, 51.2%; group 2B, 64.7%, 80.4%, 44.2%; group 3, 73.6%, 88.73%, 46.2%; group 4A, 84.1%, 83.3%, 84.6%; and group 4B, 61.9%, 37%, 76.9%.

Discussion

Trait Variation in the Typically Developing Population Informs Heterogeneity in ADHD. In this report we used graph theoretical tools to clarify a portion of the heterogeneity that exists within ADHD and typically developing control populations. Two important principles were identified that have the potential to advance our understanding of typical development and developmental neuropsychiatric disorders. The first tenet suggests that, on the basis of neuropsychological performance, typically developing children can be classified into distinct subgroups with high precision. The second tenet proposes that the heterogeneity in individuals with ADHD appears to be "nested" in this normal variation. As illustrated by our single subject classification procedures, comparing typically developing children with ADHD children within each of these distinct subgroups increases diagnostic accuracy (i.e., ADHD vs. non-ADHD classification) on the basis of the neuropsychological measures. This work highlights that illumination of such subgroups could potentially have significant practical importance for understanding the nature of typical development and identifying the etiologic underpinnings of complex disorders such as ADHD.

What Is the Role of Behavioral Variation in the Typically Developing Population and How Might It Arise? For years evolutionary psychologists have argued that human behavior (and that of other animal species) is under the same selective pressures as the physical traits so elegantly described by Darwin (34). Indeed, Darwin himself predicted this likeness to be the case at the end of his work *On the Origin of Species*, noting "In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation" (Darwin, 1859) (ref. 19, p. 399). As such, there are evolutionary arguments to be made with regard to how neuropsychological diversity might arise in the population.

The modeling literature related to adaptive complex social systems also provides significant evidence that suggests psychological heterogeneity is an important means by which to improve the robustness of the collective when faced with shifting environmental demands (35). It is a key feature in the stability of complex social systems and thus has likely been an important attribute of our evolving species.

With respect to our findings it should be noted that at times heritable diversity might form along a continuous dimension (i.e., unimodal distribution), whereas at others times it may form as multiple discrete strategies (i.e., multimodal distributions) (21, 36–38). In the latter proposal, average fitness would be about equal across the normal range of any given behavioral strategy, but individuals of different strategies might vary in the way they achieve fitness. Rapid and sizeable changes in environmental demands across time may have served as the driving force toward multiple "peaks" with regard to neurocognitive strategies or profiles in typical populations (21, 36), as found here.

For example, in the cognitive neuroscience literature some have recently argued that the single-nucleotide polymorphism (val158-met) of the catechol O-methyltransferase (COMT) gene, which codes for an enzyme that degrades dopamine in prefrontal cortex, may relate to evolutionary trade-offs between efficient executive functioning (met) and improved emotional regulation (val) (21,

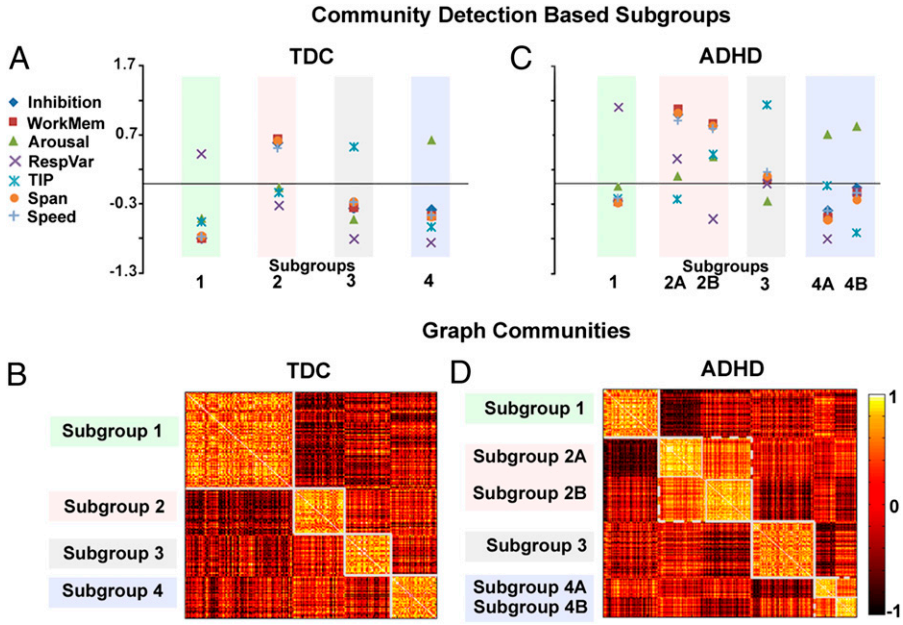

Fig. 4. Community detection identified subgroups. (*A*) After applying the community detection procedure to the typically developing cohort, four unique subgroups (i.e., cognitive profiles) emerge (*y* axis = *z* score). The community structure is depicted by correlation matrices shown in *B*. These correlation matrices represent a 213 × 213 matrix (for TDC) and 285 × 285 matrix (for ADHD). On the grid, darker colors reveal lower or negative correlations between subjects, and lighter colors reveal positive correlations between subjects. Identified communities are outlined in white. (*C*) Independently applying the community detection algorithm to the ADHD cohort shows similar findings to those in *A*. The difference between the two appears to be split in subgroup 2 and subgroup 4. The correlation matrices of the ADHD cohort are presented in *D*.

37). In this proposal each allele is maintained in the population because each provides an environment-specific selective advantage—one for cognitive efficiency under typical conditions and the other for emotional resilience under stressful circumstances (21, 37). Similar arguments have been made with regard to COMT and working memory, with an evolutionary trade-off between efficient working memory *updating* (val) and robust working memory *maintenance* (met) (22). It might be reasoned that the discrete subgroups observed here (Figs. 3 and 4) are a small representation of similar forms of neuropsychological diversity.

Just as important, these data suggest that heterogeneity within developmental neuropsychiatric disorders, such as ADHD, might be nested within this normal variation. This proposal provides a unique, and perhaps fruitful, way to conceptualize and study heterogeneity in developmental neuropsychiatric disorders.

Atypical Neuropsychological Patterns in ADHD Are Specific to Subgroup Membership. In general all of the measures examined here have been previously reported as atypical in ADHD (ref. 13 and Fig. 3). The question then is, Are these measures equally atypical within all of the identified subgroups?

For the purposes of simplicity, after the community detection analysis, we labeled each identified subgroup on the basis of the "standout" factor(s) that was lower relative to the general cohort for the TDC population. This procedure left us with four groups, which we label as follows: a variability group (subgroup 1), a low executive group (subgroup 2), a low temporal information processing group (subgroup 3), and a low arousal group (subgroup 4).

Interestingly, independently applying Newman's community detection algorithm (29) on the ADHD cohort replicated this finding, showing similar subgroup patterns (Fig. 4). The main deviation from the finding in TDC was the revelation of an additional low arousal group (subgroup 4B) and an additional low executive group (subgroup 2B). This particular result, again, suggests that some of the heterogeneity within ADHD appears to be nested within the variation found in typically developing populations. If we begin to examine the cognitive deficits within these subgroups (Fig. 3 and Table S3), rather than comparing across the entire cohorts of ADHD and TDC (Fig. 3 and Table S3), we see unique defining patterns in the deficits based on specific profiles.

For example, poor response inhibition has been suggested to be a critical component of ADHD (39) and has been well established at the group level (40). Our initial analysis comparing the entire ADHD cohort to the TDC population replicates this particular finding (Fig. 3). However, a closer look at this effect based on the subgroups identified in our graph analysis shows that this measure is atypical in only three of the six ADHD subgroups relative to controls within the same profile.

A similar finding is reported for two other well-known deficits in ADHD—working memory and temporal information processing. Again, whereas the comparison of the full ADHD and TDC cohort replicated previous findings in this regard (41), the comparisons based on community structure yielded only a subset of profiles with working memory and temporal information processing deficits. Qualitatively stronger results appeared for response variability, which was atypical in four of the six communities, supporting theories of its importance in ADHD (8). Finally, weaker, although statistically significant, contributions toward ADHD status were also identified for spatial span and speed, but again in only a subset of the identified profiles (Fig. 3 and Table S3). These results support the claim outlined in the introductory section of this paper: Whereas numerous unique neuropsychological measures have been proposed as related to ADHD (13), each of them appears to apply to only a subset of those with the disorder.

Individual Classification Based on Neuropsychological Measures Improves After Community-Based Subtyping. We note that the robustness of the identified communities was tested in several ways (e.g., Figs. S2 and S3), indicating that the group assignments identified here are highly deviant from random. One of these tests was the single-subject classification using the SVM-based MVPA. Here we demonstrated that there is enough information in the applied cognitive battery to make valid predictions of community assignment for individual subjects. In addition, we see that diagnostic classification of ADHD status for individuals in most of the subgroups, albeit modest in some, can be identified with higher accuracy when the community assignments were taken into account. The ability to characterize individual subjects and identify ADHD status empirically on the basis of a cognitive battery opens the door for tractable genetic, functional, and clinical applications. However, a crucial next step will be to evaluate the temporal stability of these cognitive types to see whether they improve on the temporal instability of DSM-IV clinical types.

Current Approach Can Be Extended Across Multiple Modalities for Other Neuropsychiatric Disorders. Although we believe our reported findings are quite provocative, we do not claim that they are exhaustive. Repeating our analysis using additional or unique neuropsychological domains might further classify or characterize profiles. Likewise, extending the present approach to neuroimaging studies may accelerate such characterization and inform heterogeneity in the neurobiology underlying the typical developing and ADHD populations (42, 43). This approach would allow for a more streamlined procedure to help investigators target and test theories related to multiple unique pathways or circuits related to ADHD (42, 43). Indeed, the same methods may be well suited to categorize participants on the basis of the neural circuitry itself or genetic markers (or pathways) rather than neuropsychological domains. We hypothesize that each group identified here is represented by distinct, multiple brain profiles. Finally we note that the approach used here is not limited to the study of ADHD, but could be used to inform heterogeneity and perhaps etiology in other developmental or adult neuropsychiatric disorders, and thus has broad appeal.

Methods

Subjects and Demographics. Children who completed a full research psychiatric evaluation aged 6–17 y participated in this study (TDC, n = 213; ADHD, n = 285) (*SI Text*). Demographic details are listed in Table S1.

Background Measures of Cognitive Functioning. Youth completed a diagnostic screening along with other testing, which included a short form of the *Wechsler Intelligence Scale for Children*, Fourth Edition (WISC-IV) (44). The age-adjusted standardized score was used as the estimate of full-scale IQ.

Neuropsychological Measures Theorized to Relate to ADHD. The neuropsychological battery was designed to capture working memory (41), response inhibition (39, 45), response variability (8), temporal information processing (46), arousal and activation (47), interference control (48), and response speed (49). All of our measures are listed in Table S2. Detailed explanations of each measure are provided in *SI Text*.

Factor Analysis Data Reduction for Neuropsychological Measures. Rational data reduction of our measures was accomplished via confirmatory factor analysis (Fig. 2). Fit for all of our models was evaluated using several indexes as noted in *Results* and *SI Text*, using the latent variable modeling program, MPLUS.

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. This procedure created two square correlation matrices providing distance information (i.e., a correlation) between any given subject pair within the ADHD and TDC cohorts (Fig. 4). Subsequent community detection (29) was applied to these matrices separately. The threshold for connected vs. unconnected pairs in each cohort was based on the maximum threshold where *reachability* remained equal to 1 (*SI Text*). This reachability threshold for the TDC graph was at r = 0.56, and the threshold for the ADHD graph was r = 0.73. 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 threshold selection (50). Both additional procedures yielded largely consistent results (Fig. S2 and *SI Text*). The strength of the modularity assignments was based on the quality index (*Q*), VOI (29, 51), and simulations created by iteratively repeating our analyses after randomizing the factor scores across participants (*SI Text* and Figs. S2 and S3). All of the preceding calculations were performed in MATLAB (Mathworks), using scripts generously provided by Olaf Sporns, Mikail Rubinov, and other collaborators (52) (Indiana University, Bloomington, IN).

Support Vector Machine-Based Multivariate Pattern Analysis. For the SVM-based MVPA we use Spider (http://people.kyb.tuebingen.mpg.de/spider/main.html), an object-orientated environment for machine learning in MATLAB. Full details of this procedure are provided in *SI Text*.

ACKNOWLEDGMENTS. We thank Bria L. Thurlow and Swathi Iyer for their helpful comments and figure preparation. We also thank the funding institutions and grants that supported this work: Grant R00 MH091238 (to D.A.F.), Oregon Clinical and Translational Research Institute (D.A.F.), Medical Research Foundation (D.A.F.), and Grants R01 MH086654 and R01 MH59105 (both to J.T.N.).

1. Hyman SE (2010) The diagnosis of mental disorders: The problem of reification. *Annu Rev Clin Psychol* 6:155–179.
2. Miller G (2010) Psychiatry. Beyond DSM: Seeking a brain-based classification of mental illness. *Science* 327:1437.
3. Agid Y, et al. (2007) How can drug discovery for psychiatric disorders be improved? *Nat Rev Drug Discov* 6:189–201.
4. Pelham WE, Foster EM, Robb JA (2007) The economic impact of attention-deficit/hyperactivity disorder in children and adolescents. *J Pediatr Psychol* 32:711–727.
5. Boyle CA, et al. (2011) Trends in the prevalence of developmental disabilities in US children, 1997-2008. *Pediatrics* 127:1034–1042.
6. Wakefield JC (1997) Diagnosing DSM-IV—Part I: DSM-IV and the concept of disorder. *Behav Res Ther* 35:633–649.
7. Berger A, Posner MI (2000) Pathologies of brain attentional networks. *Neurosci Biobehav Rev* 24:3–5.
8. Castellanos FX, Tannock R (2002) Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nat Rev Neurosci* 3:617–628.
9. Sonuga-Barke EJ, Sergeant JA, Nigg J, Willcutt E (2008) Executive dysfunction and delay aversion in attention deficit hyperactivity disorder: Nosologic and diagnostic implications. *Child Adolesc Psychiatr Clin N Am* 17:367–384, ix.
10. Sonuga-Barke EJ (2002) Psychological heterogeneity in AD/HD—a dual pathway model of behaviour and cognition. *Behav Brain Res* 130:29–36.
11. Sonuga-Barke EJ (2005) Causal models of attention-deficit/hyperactivity disorder: From common simple deficits to multiple developmental pathways. *Biol Psychiatry* 57:1231–1238.
12. Nigg JT, Goldsmith HH, Sachek J (2004) Temperament and attention deficit hyperactivity disorder: The development of a multiple pathway model. *J Clin Child Adolesc Psychol* 33(1):42–53.
13. Nigg JT, Willcutt EG, Doyle AE, Sonuga-Barke EJ (2005) Causal heterogeneity in attention-deficit/hyperactivity disorder: Do we need neuropsychologically impaired subtypes? *Biol Psychiatry* 57:1224–1230.
14. Solanto MV, et al. (2001) The ecological validity of delay aversion and response inhibition as measures of impulsivity in AD/HD: A supplement to the NIMH multimodal treatment study of AD/HD. *J Abnorm Child Psychol* 29:215–228.
15. Sonuga-Barke E, Bitsakou P, Thompson M (2010) Beyond the dual pathway model: Evidence for the dissociation of timing, inhibitory, and delay-related impairments in attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 49:345–355.
16. Insel T, et al. (2010) Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am J Psychiatry* 167:748–751.
17. Hyman SE (2007) Can neuroscience be integrated into the DSM-V? *Nat Rev Neurosci* 8:725–732.
18. Lickliter R, Honeycutt H (2003) Developmental dynamics: Toward a biologically plausible evolutionary psychology. *Psychol Bull* 129:819–835.
19. Buss DM, Reeve HK (2003) Evolutionary psychology and developmental dynamics: Comment on Lickliter and Honeycutt (2003). *Psychol Bull* 129:848–853.
20. Buss DM (1991) Evolutionary personality psychology. *Annu Rev Psychol* 42:459–491.
21. Goldman D, Oroszi G, Ducci F (2005) The genetics of addictions: Uncovering the genes. *Nat Rev Genet* 6:521–532.
22. Braver TS, Cole MW, Yarkoni T (2010) Vive les differences! Individual variation in neural mechanisms of executive control. *Curr Opin Neurobiol* 20:242–250.
23. Chapman BP, Goldberg LR (2011) Replicability and 40-year predictive power of childhood ARC types. *J Pers Soc Psychol* 101:593–606.
24. Todd RD, et al. (2001) Familiality and heritability of subtypes of attention deficit hyperactivity disorder in a population sample of adolescent female twins. *Am J Psychiatry* 158:1891–1898.
25. Martel MM, Goth-Owens T, Martinez-Torteya C, Nigg JT (2010) A person-centered personality approach to heterogeneity in attention-deficit/hyperactivity disorder (ADHD). *J Abnorm Psychol* 119:186–196.
26. Jester JM, et al. (2008) Trajectories of childhood aggression and inattention/hyperactivity: Differential effects on substance abuse in adolescence. *J Am Acad Child Adolesc Psychiatry* 47:1158–1165.
27. Lubke GH, et al. (2007) Subtypes versus severity differences in attention-deficit/hyperactivity disorder in the Northern Finnish Birth Cohort. *J Am Acad Child Adolesc Psychiatry* 46:1584–1593.
28. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10:186–198.
29. Newman ME (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103:8577–8582.
30. Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430.
31. Rizk-Jackson A, et al. (2011) Evaluating imaging biomarkers for neurodegeneration in pre-symptomatic Huntington's disease using machine learning techniques. *Neuroimage* 56:788–796.
32. Pennington BF (1991) *The Neuropsychology of Learning Disabilities* (Guilford, New York).
33. Nigg JT (2005) Neuropsychologic theory and findings in attention-deficit/hyperactivity disorder: The state of the field and salient challenges for the coming decade. *Biol Psychiatry* 57:1424–1435.
34. Darwin C (1859) *On the Origin of Species by Means of Natural selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London).
35. Miller JH, Page SE (2007) *Complex Adaptive Systems: An Introduction to Computational Models of Social Life* (Princeton Univ Press, Princeton).
36. Nettle D (2006) The evolution of personality variation in humans and other animals. *Am Psychol* 61:622–631.
37. Mier D, Kirsch P, Meyer-Lindenberg A (2010) Neural substrates of pleiotropic action of genetic variation in COMT: A meta-analysis. *Mol Psychiatry* 15:918–927.
38. Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat Neurosci* 12:1062–1068.
39. Barkley RA (1997) Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychol Bull* 121:65–94.
40. Willcutt EG, et al. (2010) Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. *Cortex* 46(10):1345–1361.
41. 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.
42. Casey BJ, Nigg JT, Durston S (2007) New potential leads in the biology and treatment of attention deficit-hyperactivity disorder. *Curr Opin Neurol* 20:119–124.
43. Durston S, van Belle J, de Zeeuw P (2011) Differentiating frontostriatal and fronto-cerebellar circuits in attention-deficit/hyperactivity disorder. *Biol Psychiatry* 69:1178–1184.
44. Wechsler D (2003) *Wechsler Intelligence Scale for Children Technical and Interpretive Manual* (The Psychological Corporation, San Antonio), 4th Ed.
45. 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.
46. Toplak ME, Dockstader C, Tannock R (2006) Temporal information processing in ADHD: Findings to date and new methods. *J Neurosci Methods* 151:15–29.
47. Sergeant JA, Oosterlaan J, van der Meere JJ (1999) *Information Processing and Energetic Factors in Attention Deficit/Hyperactivity Disorder* (Kluwer/Plenum, New York).
48. Welsh MC, Pennington BF (1988) Assessing frontal lobe functioning in children: Views from developmental psychology. *Dev Neuropsychol* 4:199–230.
49. 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.
50. Rubinov M, Sporns O (2011) Weight-conserving characterization of complex functional brain networks. *Neuroimage* 56:2068–2079.
51. 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.
52. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 52:1059–1069.

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

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

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, $\boldsymbol{w}$ 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

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

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

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}\right).$$

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.

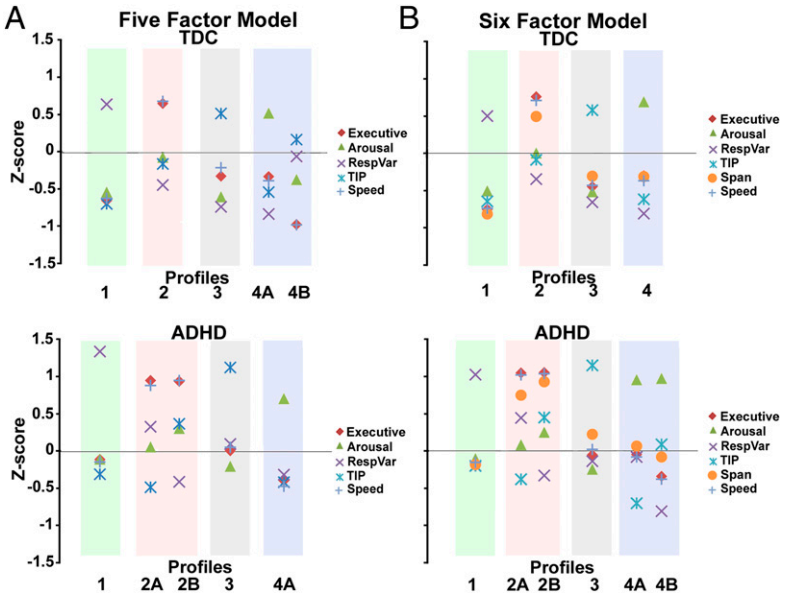

Fig. S1. Community detection applied to alternative five- and six-factor models. As noted in the main text because a limitation of confirmatory factor analysis is the potential for untested models to have equivalent fit, we tested several competing models for our feature reduction step. (*A*) Community 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.

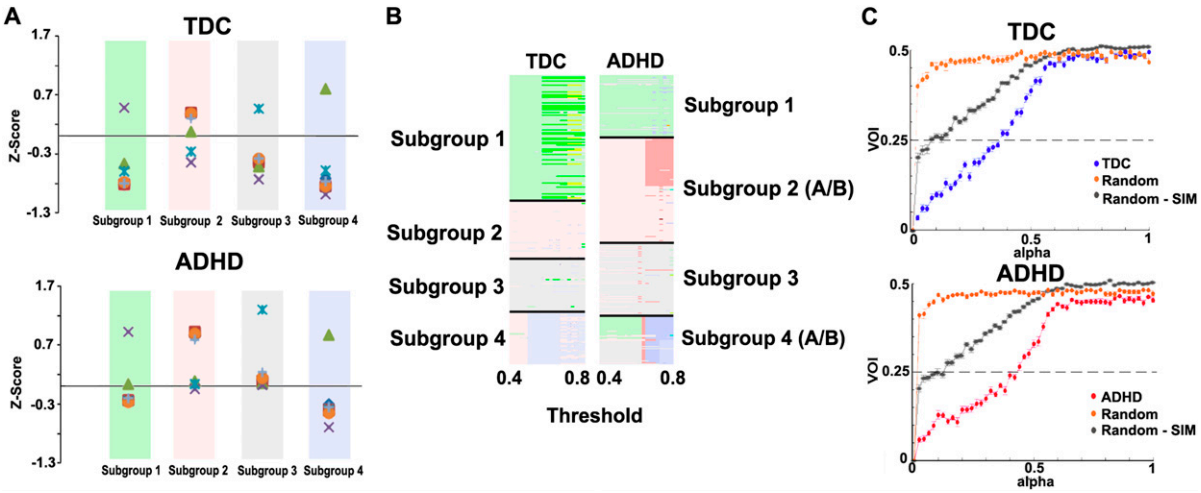

Fig. S2. (*A*) To ensure our subgroups in both the ADHD and the TDC populations were not specific to a particular threshold, we reexamined our data using a weight-preserving modularity algorithm (23). Our subgroupings in both the TDC and the ADHD populations replicated using this procedure. *Q* values 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.

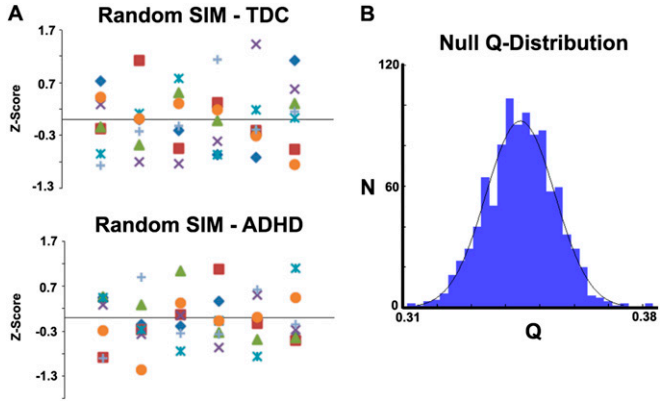

Fig. S3. Generating a null distribution of *Q* on the basis of the weight-preserving modularity algorithm introduced by Rubinov and Sporns (23). (*A*) One example of our simulations created by repeating our analyses after 1,000 iterations of randomizing the factor scores across participants. Whereas 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.

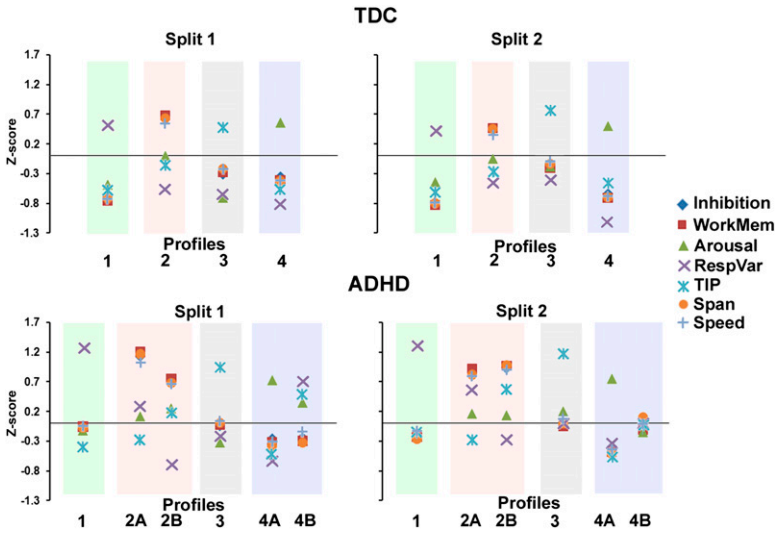

Fig. S4. A split replication procedure was applied for the SVM-based MVPA analysis. To use our SVM to test the ability for single-subject classification into any given community we split the data into a test and a training set. Community detection was applied to each split separately and revealed similar 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).