A low-level perceptual correlate of behavioral and clinical deficits in ADHD

Andra Mihali[1,2,7], Allison G Young[3,7], Lenard A. Adler[3], Michael M. Halassa[1,3,4,5,6], Wei Ji Ma[1,2]

1 Center for Neural Science, New York University, New York, NY, USA

2 Department of Psychology, New York University, New York, NY, USA

3 Department of Psychiatry, NYU School of Medicine, New York, NY, USA

4 Neuroscience Institute, NYU School of Medicine, New York, NY, USA

5 Department of Neuroscience and Physiology, NYU School of Medicine, New York, NY, USA

6 Department of Brain and Cognitive Science, MIT, Boston, MA, USA

7 Andra Mihali and Allison G Young contributed equally to this work.

Corresponding authors are Wei Ji Ma (weijima@nyu.edu) and Michael M Halassa (mhalassa@mit.edu).

In many studies of attention-deficit hyperactivity disorder (ADHD), stimulus encoding and processing (perceptual function) and response selection (executive function) have been intertwined. To dissociate these functions, we introduced a task that parametrically varied low-level stimulus features (orientation and color) for fine-grained analysis of perceptual function, and that also required participants to switch their attention between feature dimensions on a trial-by-trial basis, thus taxing executive processes. Our response paradigm captured task-irrelevant motor output (TIMO), reflecting failures to use the correct stimulus-response rule. ADHD patients had substantially worse perceptual function than Controls, especially for orientation. ADHD participants had also higher TIMO; this measure was strongly affected by the switch manipulation. Across participants, the perceptual variability parameter was correlated with TIMO, suggesting that perceptual deficits could underlie executive function deficits. Based on perceptual variability alone, we were able to classify participants into ADHD and Controls with a mean accuracy of about 77%. Participants' self-reported General Executive Composite score correlated not only with TIMO but also with the perceptual variability parameter. Our results highlight the role of perceptual deficits in ADHD and the usefulness of computational modeling of behavior in dissociating perceptual from executive processes.

In Attention Deficit Hyperactivity Disorder (ADHD), self-reported behavioral deficits have been attributed to differences in executive function, attention, and perceptual function. These brain functions can be objectively quantified with behavioral paradigms, but the individual components are often not well separated. ADHD patients tend to have worse executive function than Controls according to several metrics [1–4], predominantly in response execution and inhibition [5–7], but also in switching between stimulus-response rules [8–11]. It has been pointed out that non-executive processes, such as alertness, accumulation of evidence or drift rate, motivation, or reward processing, might interact with executive ones [12–16]. In the realm of visual attention, differences in accuracy or reaction time have been found in some visual search tasks but not in others [17,18], and no consistent deficits have been found in visuo-spatial orienting [19–22]. Despite evidence of impaired perceptual function in ADHD from psychiatric assessments [23,24], behavioral studies that examined the quality of perceptual encoding in ADHD in the absence of executive or attentional involvement found small and inconsistent differences (see [25] for a review). There have been attempts to dissociate executive from perceptual function in ADHD within a single task, but the results have been mixed [26–29]. One complicating factor could be that commonly used stimuli such as faces, letters, or numbers are high-dimensional and have content at many levels. Another complication is that if the observer uses only 2 response keys in a task-switching paradigm, an error could be

either due to a failure to switch or to a successful switch followed by a perceptual error.

Here, we attempted to dissociate perceptual from executive deficits in ADHD using a visuo-motor decision-making paradigm with task-switching which avoids the complications listed above. Our response paradigm allowed for *task-irrelevant motor output* (TIMO), a new measure of executive control deficits. Moreover, we optimized the quantitative characterization of perceptual deficits by: a) using simple stimuli with feature dimensions orientation and color, thus minimizing high-level cognitive effects; b) varying stimuli parametrically along a continuum to estimate psychometric curve parameters (standard in perceptual psychophysics but still relatively rare in the study of ADHD [22,26–28,30]); c) using an efficient stimulus selection method to minimize the number of trials needed for accurate estimation [31]. Our work follows a recent proposal to apply four levels of analysis to computational psychiatry: development of behavioral tasks, fitting of computational models, estimating parameters, and classification for diagnosis [32].

20 ADHD and 20 Control adult participants took part in our experiment(see Supplementary section "Demographic and clinical information"). Stimuli were colored ellipses; each display contained one stimulus on the right of the fixation dot and one on the left. The participant either performed fine orientation discrimination (clockwise or counterclockwise) or fine color discrimination (more yellow or more blue). Every trial started with a symbolic feature dimension cue, informing the participant which feature dimension was relevant on that trial. Simultaneously presented was a spatial cue (a line segment), informing the participant which side of the screen was relevant on that trial (Figure 1a). To better detect failures of spatial or feature switching, we used a response paradigm in which, on each trial, only 2 of 8 response keys were relevant, depending on the spatial and the feature cue; any other key press counted as task-irrelevant motor output (TIMO).

Each participant experienced three types of blocks: ones in which orientation was always relevant (Ori condition), ones in which color was always relevant (Col condition), and switch blocks, in which the relevant feature dimension was randomly chosen on each trial; we will refer to the orientation and color trials in switch blocks as the OriS and ColS conditions, respectively, and to the difference between no-switch and switch blocks as a difference in (executive) load. The spatial cue was randomly chosen on each trial. Separately in each condition and for each participant, we used a Bayesian adaptive method to select maximally informative stimuli (see "Target stimulus selection" in Methods). This method allowed us to estimate the psychometric curve parameters with relatively few trials. An observer's sequence of computations in the task can be conceptualized as a perceptual decision-making stage (stimulus encoding, affected by attention, and inference), followed by executive processing (rule retrieval and response execution) (Figure 1b). The parametric variation of stimulus strength allowed us to estimate perceptual variability as a main metric of perceptual function, and the 8-button response paradigm allows us to estimate task-irrelevant motor output as a main metric of executive function. In addition, we characterized behavior using other psychometric curve parameters, median reaction time, and reaction time variability.

Since TIMO refers to the trials when participants pressed one of the 6 irrelevant keys, such responses could reflect a failure of switching or proper stimulus-response rule retrieval, but are unlikely to be affected by perceptual failures in the relevant feature dimension. TIMO was quite low overall (mean $\pm$ sem: 0.061 $\pm$ 0.007), but ADHD participants produced a higher proportion of TIMO (Figure 2a). A three-way mixed-design ANOVA on log TIMO with between-participants variable group and within-participants factors load (No-switch and Switch) and feature (Ori and Col) reveals a significant effect of group ($F(1,38) = 8.83, p = 0.005, \eta_p^2 = 0.19$), a significant effect of load ($F(1,38) = 101.4$, $p < 0.0001, \eta_p^2 = 0.73$), and no significant effect of feature ($F(1,38) = 1.62$, $p = 0.21, \eta_p^2 = 0.04$). Neither of the two-way interactions nor the three-way interaction were significant ($p > 0.06$). In particular, the group $\times$ load interaction was not significant ($F(1,38) = 3.72, p = 0.06, \eta_p^2 = 0.09$); thus, we did not find that switching between feature dimensions carries a higher cost in ADHD. Next, we performed 12 post-hoc planned comparisons: within each group, Wilcoxon signed-rank tests for Ori versus OriS, Col versus ColS, Ori versus Col, and OriS versus ColS, and between groups, Wilcoxon rank-sum tests for Ori, OriS, Col, and ColS. After Sidak correction ($\alpha = 0.0043$), no between-group comparisons were significant ($p > 0.0046$). The within-group load comparisons were all significant ($p < 0.002$). No within-group feature comparisons were significant ($p > 0.07$). In the OriS and ColS conditions, the majority of TIMO seemed to be feature errors (Figure S1b). Taken together, these results validate TIMO as a metric of interest for executive control.

ADHD participants seemed to have longer reaction times (RTs; Figure 2b). Indeed, three-way mixed-design

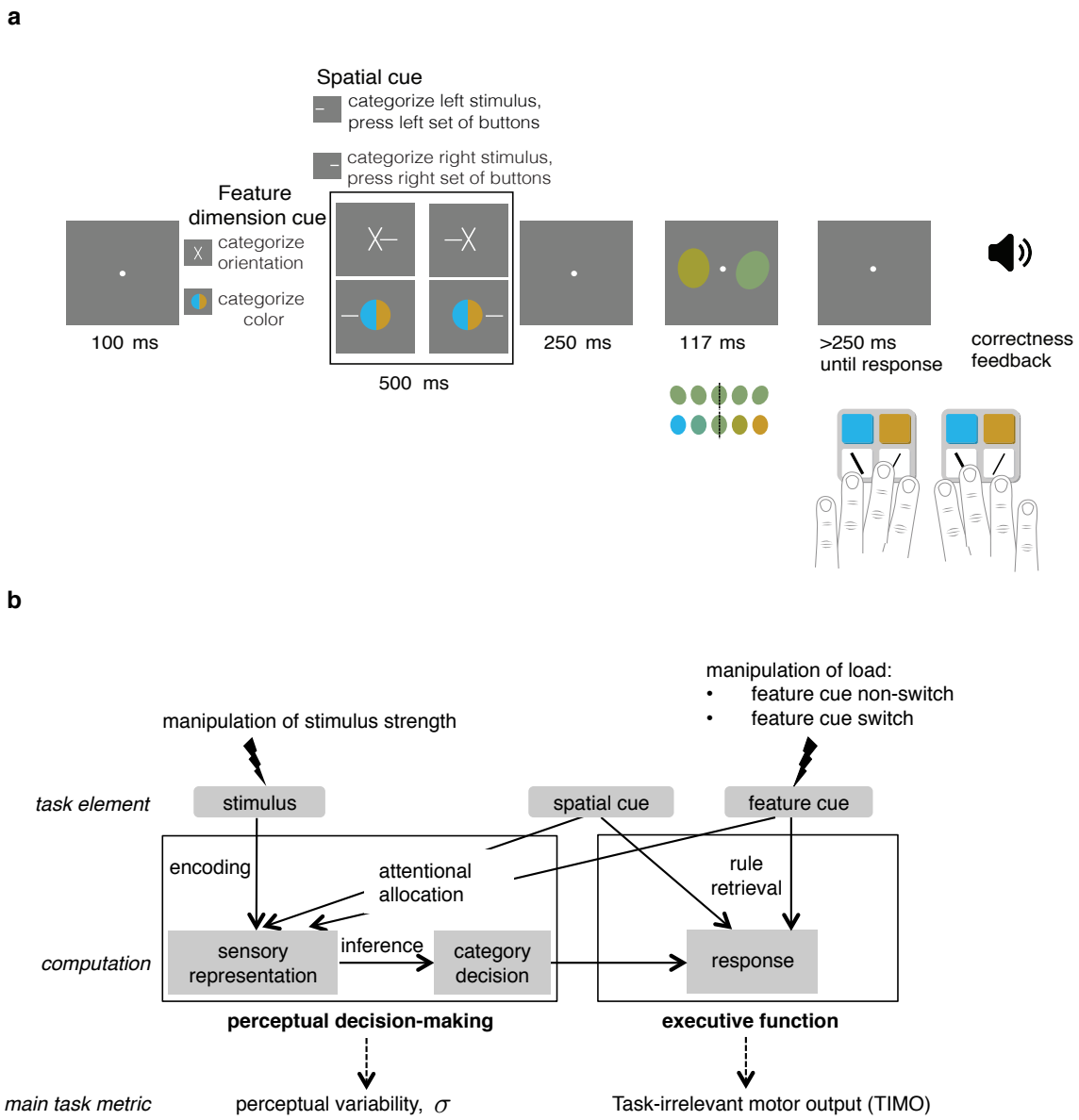

Figure 1: **Task design and dissociation of perceptual and executive processes.** **(a)** Trial sequence. A feature dimension cue indicated whether orientation (cross) or color (colored circle) was relevant, while a simultaneous endogenous spatial cue (line segment) indicated which side (left or right) was relevant. Thus, the participant received one of 4 possible cue screens. We always chose the spatial cue randomly. The participant had to respond whether the orientation of the ellipse on the relevant side was clockwise or counterclockwise with respect to vertical, or whether its color was more yellow or more blue, with the associated set of keys (left or right). The participant could press any one of 8 keys but only 2 were task-relevant on a given trial. The participant received correctness feedback. For full details, see Methods. **(b)** Schematic of the perceptual and executive processes that may play a role in this task, and the corresponding task metrics.

ANOVA on log RTs revealed a significant effect of group ($F(1,38) = 4.72, p = 0.036, \eta_p^2 = 0.11$), and significant effects of load ($F(1,38) = 84.92, p < 0.0001, \eta_p^2 = 0.69$) and feature ($F(1,38) = 70.23, p < 0.0001, \eta_p^2 = 0.65$). In addition, we found significant group × feature ($F(1,38) = 4.63, p = 0.038, \eta_p^2 = 0.11$) and load × feature interactions ($F(1,38) = 12.37, p = 0.001, \eta_p^2 = 0.25$), but not a significant group × load interaction ($F(1,38) = 3.97, p = 0.054, \eta_p^2 = 0.095$). After Sidak correction ($\alpha = 0.0043$), none of the between-group comparisons were significant ($p > 0.019$). The effects of within-group load and feature on log RTs were all significant both within Control and within ADHD ($p < 0.001$). Higher RTs for Col than Ori could be due to the fact that the Ori responses are intuitively mapped to left and right, while the Col responses are arbitrarily mapped as blue to left and yellow to right. Higher RT variability (or intra-individual variability) in ADHD has been found consistently [4] and has been generally thought to reflect cognitive processes separate from higher median RTs [4, 33] (but see [34] for an opposing account). The term RT variability has been used to refer to different aspects of RT distributions [4]; here, we fitted ex-Gaussian distributions [35] and used the τ parameter as a measure of RT variability. The τ parameter has been shown to capture the tendency of ADHD participants to have a higher proportion of abnormally slow responses [4, 33, 35]. Before committing to the ex-Gaussian, we verified that it captures the data better than the log-normal and gamma distributions (see Supplementary section "Further information on reaction times"). Three-way mixed-design ANOVA on $\log \tau$ revealed a significant effect of group ($F(1,38) = 7.72, p = 0.008, \eta_p^2 = 0.17$), an effect of load ($F(1,38) = 9.32, p = 0.004, \eta_p^2 = 0.20$) and an effect of feature ($F(1,38) = 18.85, p < 0.001, \eta_p^2 = 0.33$). The only significant interaction was between load and feature ($F(1,38) = 14.96, p < 0.001, \eta_p^2 = 0.28$). After Sidak correction ($\alpha = 0.0043$), none of the between-group comparisons were significant ($p > 0.006$). Within Controls, the effects of load and feature on log RT τ were significant for Ori vs. OriS and Ori vs. Col ($p < 0.001$). Within ADHD, no effects of load or feature were significant ($p > 0.02$). We confirmed the pattern of higher RT variability in ADHD with a non-parametric measure, RT iqr (see Supplementary section "Further information on reaction times"). Overall, we found that ADHD participants had longer and more variable reaction times, consistent with previous work [4, 35, 36]. However, RT-related differences are difficult to categorize as perceptual or executive because they might encompass multiple processes, including sensory encoding (possibly slower with more perceptual variability), decision time, stimulus-response rule retrieval, response preparation, and response execution [34]. The effect of load on RT and RT τ does seem to suggest that on Switch trials, more time is spent on executive processes, here stimulus-response rule retrieval, response preparation, and response execution, relative to No-Switch trials.

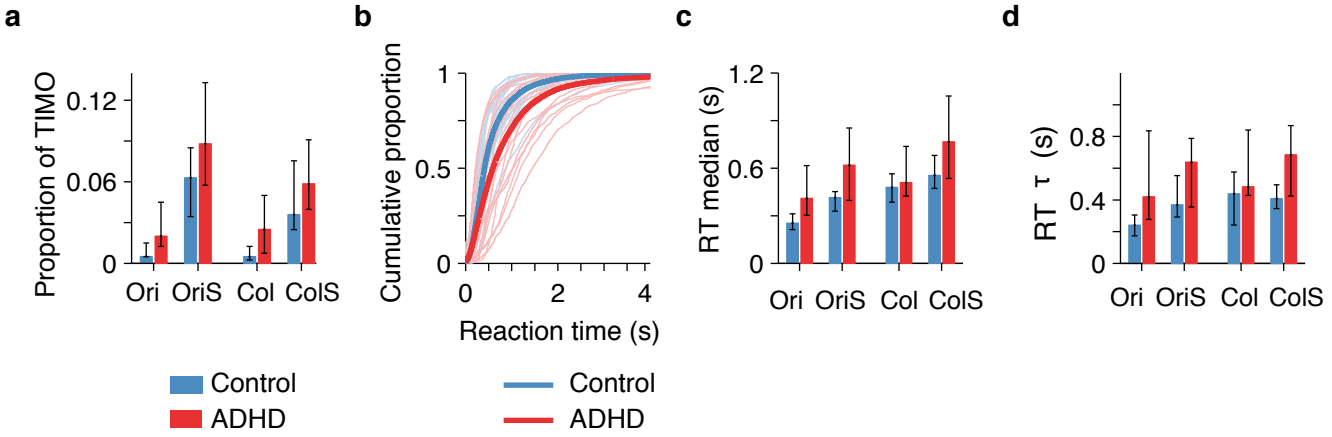

Figure 2: **ADHD participants had higher TIMO and longer and more variable reaction times. (a)** Proportion of TIMO across conditions. Here and elsewhere, values represent medians across participants and error bars the bootstrapped 95% confidence intervals. **(b)** Empirical cumulative density functions of reaction times, collapsed across all conditions. Thin lines: individual participants. Thick lines: median for the RT distribution collapsed across all participants in a group. **(c)** Reaction time median by condition and group. Throughout the paper, we use RT median because reaction time distributions are not Gaussian. **(d)** Reaction time variability metric, the τ parameter from ex-Gaussian distribution fits, by condition and group.

We confined the following analysis to the trials in which participants pressed one of the 2 relevant keys. Because of the Bayesian stimulus selection method, proportion correct is largely stable across conditions and participants (mean $\pm$ sem: 0.811 ± 0.005, Figure S1A) and thus not informative. Instead, we fitted a psychometric curve to non-TIMO trials in each condition [37]. The normalized orientation and color continua spanned the interval $[-0.5, 0.5]$. Each psychometric curve has three parameters: a point of subjective equality μ, perceptual variability σ, and a lapse rate λ (Figure 3b,c and Figure S6d). Non-zero μ represents a tendency to choose one option more than the other. The parameter σ is a composite of sensory noise and decision noise, and also reflects the quality of the allocation of spatial attention, and of feature attention in switch blocks. The parameter λ reflects trials with lapses in attention or erroneous motor output. In our main model, we assumed that μ and λ are independent of load; we confirmed this assumption by comparing to a model without these assumptions ("full" model) in the Supplementary section "Psychometric curve parameters". The parameters σ and λ might trade off against each other, although this is less of a concern in our main model than in the "full" model. Three-way mixed-design ANOVA on log σ showed a significant effect of group ($F(1, 38) = 10.56, p = 0.002, \eta_p^2 = 0.22$), a significant effect of feature ($F(1, 38) = 37.2, p < 0.001, \eta_p^2 = 0.49$), but not of load ($F(1, 38) = 0.97, p = 0.33, \eta_p^2 = 0.025$). The effect of group $\times$ load was not significant ($F(1, 38) = 2.0, p = 0.16, \eta_p^2 = 0.05$). Because the normalization to the (arbitrary) stimulus range is specific to each feature dimension, the values of σ cannot be meaningfully compared between orientation and color: a different stimulus range would have changed the σ values without changing the observer's true perceptual variability. Therefore, only the within-feature post-hoc comparisons are meaningful, giving a corrected significance level of $\alpha = 0.0065$. Then, the between-group comparisons were significant for both Ori and OriS ($p < 0.0005$), but not for Col ($p = 0.0083$) or ColS ($p = 0.28$). No post-hoc comparisons with load were significant neither within Controls nor within ADHD ($p > 0.01$). Higher σ for orientation in ADHD could result from worse low-level sensory encoding (e.g. higher neural noise), lower covert endogenous attention, higher decision noise, or even noise in the inference process about the perceptual category. The lapse rate reflects responses that are independent of the stimulus, such as lapses of attention, but could also trade off with the σ parameter. Two-way mixed-design ANOVA on log λ showed a large effect of feature ($F(1, 38) = 28.08, p < 0.0001, \eta_p^2 = 0.43$), but no significant effect of group ($F(1, 38) = 1.72,\ p = 0.19, \eta_p^2 = 0.04$) and no significant group $\times$ feature interaction. After Sidak correction ($\alpha = 0.0127$), we found that Control ($p < 0.0001$) and ADHD ($p = 0.011$) participants tended to lapse more for color than for orientation, possibly because the stimulus-response mapping is less intuitive. Results for μ are in the Supplementary section "Psychometric curve parameters". In conclusion, the parametric variation of low-level stimulus variables combined with stimulus optimization revealed robust perceptual deficits in ADHD, specifically for orientation.

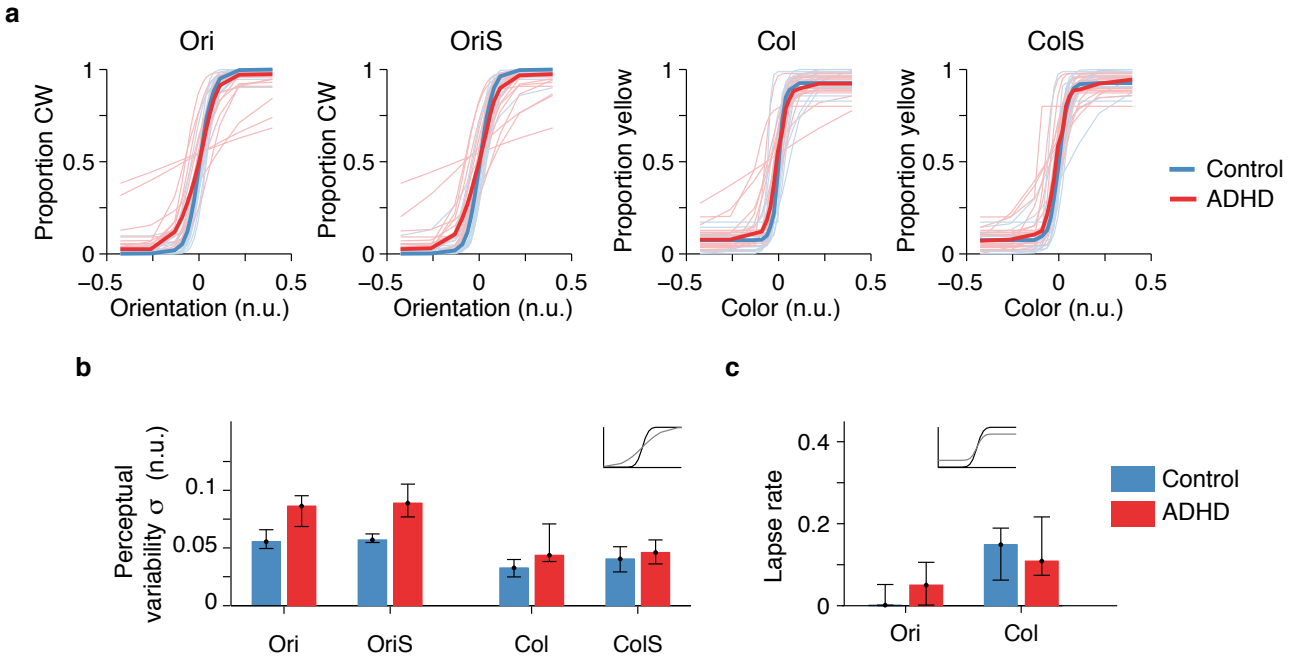

Figure 3: **Fitted psychometric curves and parameters. ADHD participants had higher perceptual variability.** **(a)** Psychometric curve fits across all conditions. Here and elsewhere, n.u. stands for normalized units. Thin lines: individual participants. Thick lines: medians for each group. For fits overlaid on top of data, see Figure S6. **(b)** Perceptual variability parameter values, medians and bootstrapped 95% confidence intervals. Top inset plot: black psychometric curve has low noise, while the grey has higher noise. **(c)** Lapse rate. Top inset plot: black psychometric curve has low lapse, while the grey has higher lapse.

Table 1: **Pairwise Spearman correlations across task metrics (both behavioral and clinical). Both TIMO and perceptual variability are significantly correlated with several other variables.** Bold-faced denotes significance after multiple-comparisons correction ($\alpha = 0.0089$, see Methods).

	TIMO	RT	RT τ	Perceptual variability (σ)	Lapse rate (λ)	GEC
TIMO						
RT	$\boldsymbol{r}$**=0.46** $p = 0.003$					
RT τ	$\boldsymbol{r}$**=0.42** $p = 0.007$	$\boldsymbol{r}$**=0.84** $p < 0.0001$				
Perceptual variability (σ)	$\boldsymbol{r}$**=0.41** $p = 0.0085$	$\boldsymbol{r}$**=0.55** $p = 0.0003$	$\boldsymbol{r}$**=0.57** $p = 0.0002$			
Lapse rate (λ)	$\boldsymbol{r}$**=0.46** $p = 0.003$	$r = 0.23$ $p = 0.15$	$r = 0.17$ $p = 0.30$	$r = 0.28$ $p = 0.08$		
GEC	$\boldsymbol{r}$**=0.53** $p = 0.0005$	$r = 0.25$ $p = 0.12$	$r = 0.34$ $p = 0.03$	$\boldsymbol{r}$**=0.50** $p = 0.0009$	$r = 0.30$ $p = 0.06$	
ACDS	$r = 0.40$ $p = 0.01$	$r = 0.31$ $p = 0.05$	$\boldsymbol{r}$**=0.45** $p = 0.004$	$\boldsymbol{r}$**=0.51** $p = 0.0008$	$r = 0.18$ $p = 0.26$	$\boldsymbol{r}$**=0.80** $p < 0.0001$

Next, we asked whether behavioral metrics are correlated with each other (Table 1). For this analysis, we collapsed across groups; per participant, we averaged each behavioral metric across all four conditions. We found that the perceptual variability parameter σ is significantly correlated with TIMO, RT median, and RT τ, with high effect sizes. Note that the perceptual variability parameter and TIMO were computed from different sets of trials, therefore reducing the probability that their correlation is spurious. In addition, a breakdown of some of these correlations by group and condition is presented in Supplementary section "Breakdown of correlations".

So far, we have characterized behavioral differences between ADHD and Controls in our task. Next, we asked if behavioral metrics relate to common clinical ones, namely the General Executive Composite score (GEC), as assessed by the Brief-A questionnaire (self-reported, [38]), as well as the (ACDS) scores (clinician interview, [39]). The GEC and ACDS (Table S2) are meant to be continuous measures of executive control and symptom severity, respectively. Both GEC and ACDS revealed strong correlations with TIMO, suggesting that TIMO could serve as a behavioral marker of executive deficits. GEC and ACDS were both also strongly correlated with perceptual variability. In addition, ACDS (but not GEC) was correlated with RT τ, which provides a graded counterpart of the robust finding of increased RT variability in ADHD [4]. A linear regression of GEC as a function of behavioral metrics ($R^2 = 0.38$) showed only TIMO as statistically significant (Table S5a), reinforcing our interpretation of TIMO as reflecting failures of executive function. A linear regression of ACDS as a function of the same metrics ($R^2 = 0.33$) only showed significance for RT τ (Table S5b), suggesting that GEC and ACDS, despite being strongly correlated (Figure S8), could capture distinct aspects of impairment [40]. These results suggest that our behavioral metrics capture to some extent the same processes as clinical metrics, while having the advantage of avoiding the potential subjectivity inherent in questionnaires.

Finally, we asked how accurately we can classify a given participant as ADHD or Control based purely on behavioral task metrics. A logistic regression using only the perceptual variability parameter yielded a classification accuracy of 77.5% (Figure S10a). Adding TIMO as a regressor improved the accuracy to 82.5%, but adding more regressors did not yield further improvement. In order to assess the use of this logistic classification for out-of-sample prediction, we did stratified 10-fold cross-validated and found mean accuracies of 77.1% with perceptual variability as the only regressor and 77.8% with both perceptual variability and TIMO

(for more details, see Supplementary section "Prediction of clinical variables"). The relatively high classification performance suggests that our task has potential as a diagnostic tool.

In this study, we dissociated perceptual from executive deficits in ADHD with a new visuo-motor decision-making task with a task-switching component. Some previous task-switching studies used only 2 response keys, thus conflating failures to switch between rules with errors within the correct rule due to perceptual or attentional failures (i.e., [10]). To better isolate executive deficits, we used 8 response keys, 6 of which are irrelevant on any given trial (TIMO). Our finding of higher TIMO in ADHD could be due to more spatial switching errors or more feature switching errors, but it is hard to quantify these contributions since TIMO was overall relatively low. It is conceivable that types of stimuli that require spatial integration [41–44] or cross-modal switching [45], or more complex forms of task switching [1, 2], would produce larger differences on a TIMO-like executive function measure. To estimate perceptual variability accurately, we used simple stimuli that varied continuously along one dimension. ADHD participants showed more perceptual variability in the representation of simple visual stimuli (specifically in orientation). This could reflect higher sensory noise, higher decision noise, or less effective attentional allocation. The considerable magnitude of these perceptual deficits is in line with a suggestion that such deficits might be larger when executive processes are simultaneously taxed [25, 46], as is the case in our task.

In line with previous work [4,35,36], we found that ADHD participants had longer and more variable reaction times. In addition, our paradigm allowed us to analyze correlations across individuals between reaction time metrics and other metrics. The correlation between the noise parameter σ and median reaction time is consistent with a drift-diffusion model, in which slower accumulation of evidence simultaneously leads to lower accuracy and longer reaction times. Indeed, a lot of previous work in ADHD found slower drift rates in ADHD [14, 34, 47, 48]. In another line of work, it has been shown [49] and later proposed more generally [50] that higher decision noise in ADHD could underlie behavioral variability, specifically choice variability and RT variability. The main effects of group on σ – which could comprise both sensory and decision components – and RT τ, as well as their correlation, are consistent with this proposal, while also extending it to include the possibility of an even lower-level source of variability. Future work could further disentangle the influences of sensory, decision and inference noise on behavioral variability in ADHD with a paradigm such as the one in [51].

Our results could speak to the question of non-executive components underlying measured executive control deficits, as we found a significant correlation between the perceptual variability parameter and the executive control metric TIMO. Structural equation modeling could look beyond correlations at potential causal relationships between perceptual and executive deficits, as in [52]. In particular, our results raise the possibility of a shared neural source underlying perceptual and executive function deficits, such as a lower signal-to-noise ratio in early brain areas. While ideas of lowered signal-to-noise ratio implemented through impaired dopamine and noradrenaline signaling in ADHD have been put forward before, they have been confined to cerebellar, striatal and prefrontal regions [50, 53, 54]. An earlier candidate source region is the thalamic reticular nucleus, where impaired gain control has been implicated in impaired visual-auditory attentional selection in mouse [55–57]; this might be consistent with thalamic dysfunction in human ADHD [58–60]. Another possibility is the sensory cortex, such as through less selective orientation tuning of cells in V1; this was the mechanism proposed to underlie decreased orientation discrimination with aging in monkeys [61]. To distinguish between possible loci and mechanisms, recent mouse models of ADHD will be invaluable [62, 63]. More broadly, using simple rather than high-level cognitive stimuli has the advantage that they can be used in parallel humans and animals studies, and thus help shed light on the neural circuits implicated in ADHD and how medications can alter these circuits [64].

ADHD diagnosis still relies predominantly on self-reports, and widely accepted "psychomarkers" (also called "neurocognitive endophenotypes") and biomarkers are lacking [65]. For our findings to have implications for clinical practice, it is necessary that our task metrics are predictive of clinical metrics. We found that this was indeed the case. First, based on perceptual variability alone, we were able to classify participants into ADHD and Control with cross-validated mean accuracy of 77.0% (including TIMO, 77.7%). Beyond binary classification, we also found strong correlations between behavioral metrics (σ, TIMO, and RT τ) and clinical ones (GEC and ACDS). Based on these correlations, the behavioral metrics in our task could be considered candidate psychomarkers for ADHD, similar to performance on the Continuous Performance Test [66], response variability [3], and drift rate [15], and along with potential biomarkers such as microsaccade rate [67, 68],

pupil size [69], and cortical thickness [70]. Simple behavioral paradigms and modeling applied to ADHD and other disorders could in the long term help refine diagnostic categories and inform and quantify the efficacy of treatment, as is the goal in computational psychiatry more broadly [32, 71–73].

Acknowledgements

This work was supported by an NYU SoM Applied Research Support Fund (ARSF). We thank Terry Leon, Michael Silverstein, Saima Milli, Jonathan Yuh for assistance with protocol preparation, participant recruitment and clinical data management and analysis. We thank Ili Ma for useful conversations and comments on an earlier version of this manuscript. In addition, we thank Eero Simoncelli, Marisa Carrasco, Roozbeh Kiani, Mariel Roberts, and members of the Ma lab, especially Maija Honig, Luigi Acerbi, Will Adler, Bas van Opheusden and Aspen Yoo for useful conversations.

Methods

Experiment

Participants. We recruited all participants through local advertisements, including flyers and newspaper and radio advertisements. Participants in both groups were matched as much as possible by age, sex, and education (see Table S1). 20 ADHD participants (12 female) of mean age 35.3 (SD: 10.0, range: 21 to 55) and 20 control participants (11 female) of mean age 32.5 (SD: 6.1, range: 19 to 44), with no statistical difference between their ages (Wilcoxon rank-sum test, $p = 0.78$), participated. All participants spoke English and had normal or corrected-to-normal vision. We asked every participant before they started if they were colorblind. One participant was excluded because of color blindness. All participants provided informed consent. The study conformed to the Declaration of Helsinki and was approved by the Institutional Review Board of New York University School of Medicine.

Psychiatric assessment and diagnosis. None of the participants with ADHD were prescribed or took stimulant medication within 2 months of participating in the study. Participants with comorbid anxiety or unipolar depressive disorders were included as long as the symptoms at the time of evaluation were mild or in remission. Participants with bipolar disorders, psychotic disorders, substance use disorders, and neurologic disorders were excluded. For all adults, the diagnostic procedure included both clinician administered and self-administered scales. A trained clinician assessed every participant using the Adult ADHD Clinician Diagnostic Scale (ACDS) v.1.2, the Adult ADHD Investigator Symptom Rating Scale (AISRS), the Clinical Global Impressions-Severity of Illness (CGI-S) Scale, and the M.I.N.I International Neuropsychiatric Interview. All participants also completed the Adult ADHD Self-Report Scale (ASRS v.1.1.), the Adult ADHD Quality of Life (AAQoL) Scale, the World Health Organization Disability Assessment Schedule (WHODAS-II), and the Behavior Rating Inventory of Executive Function Adult Version (BRIEF-A). These scales have been extensively validated [39, 74, 75].

Apparatus. We displayed stimuli on a 23-inch (58.42 cm) Acer T232HL LCD monitor of resolution: 1920 $\times$ 1080 pixels and 60 Hz refresh rate (1 frame lasting 16.7 ms). We used a Kinesis Freestyle2 split keyboard. Participants used a head rest located at approximately 55 cm from the screen; this meant that 1 degree of visual angle (dva) subtended approximately 34.32 pixels. Stimulus presentation and response collection were controlled by a Windows computer running Matlab 7.1 (MathWorks, Massachusetts, USA) with Psychtoolbox3 [76–78] and EyeLink [79].

For 10 out of 20 ADHD participants and 10 out of 20 control participants, we monitored their fixation and recorded their eye movements. The rationale for not eye tracking all participants was a mixture of lack of sufficient time on the participants' side and balanced design on the experimenters' side. The eye tracker was calibrated using the five-point calibration routine before every block. We recorded eye movements using a remote infrared video-oculographic system (EyeLink 1000 Plus; SR Research, Ltd, Mississauga, Ontario, Canada) with a 1 kHz sampling rate. We set the heuristic filtering option OFF.

Stimuli. The background was mid-level gray (28.7 cd/m^2). The stimuli were ellipses with area of 1600 pixels^2 and 0.55° eccentricity, and thus with a major axis of 50 pixels and minor axis of 41 pixels. For the non-

target ellipse, the orientation was randomly drawn from a von Mises distribution centered at 0 with $\kappa = 30$, and then divided by 2, approximately equivalent to a Gaussian distribution with mean 0 and a standard deviation of about 5°. The color of the non-target ellipse was uniformly drawn between blue and yellow in CIE Lab color space, with blue as [78 -30 -40], corresponding in RGB space to [0 167 255] and of measured luminance 38.86 cd/m^2 and yellow as [78 0 80] corresponding in RGB to [200 130 0] and of measured luminance 39.14 cd/m^2. Lightness was always kept constant at $L = 78$. The target stimulus was specified on a trial-to-trial basis using a Bayesian algorithm (see the next subsection).

Trial sequence (Figure 1a). A trial sequence started with the simultaneous appearance of a feature dimension cue and a spatial cue, presented for 500 ms. The feature dimension cue for orientation consisted of 2 white ([250 250 250]) line segments, each of length 35.77 pixels, crossing at the center, with orientations tilted ±26.56° with respect to vertical; for color, it consisted of 2 semi-circles (divided vertically, right one yellow, left one blue) joined to form a circle of radius 10 pixels. Simultaneously, a spatial cue was presented, which consisted of a horizontal line segment of length 17.16 pixels emanating from the center of fixation to the left or to the right. We chose 500 ms to ensure sufficient time for the deployment of endogenous feature-based attention [80]. Following a delay of 250 ms consisting of the presentation of a central fixation circle of radius 4.2 pixels, 2 stimuli appeared at 2.5 dva to the right and left of a central fixation circle. The 2 stimuli (see below for information on the target stimulus) were presented for 117 ms, followed by a delay period of 250 ms.

After the delay period, the participant had to respond about the target ellipse via a specific key press out of a total of 8 keys (Figure 1a). On any given trial, 6 of these 8 keys are irrelevant. For orientation, the participants were instructed to press one of the 2 labeled keys for clockwise (CW) or counterclockwise (CCW), using the left keypad for the left spatial cue and the right one for the right spatial cue. For color, they had to press one of 2 labeled keys to indicate whether the ellipse was more yellow or more blue, also using left or respectively right keypad depending on the spatial cue. The direction of the spatial cue was randomly drawn on each trial, so participants used their right hand approximately half the time. After the response, auditory feedback was provided for 200 ms: a 1200 Hz tone if the participant had pressed the correct key, and a 500 Hz tone if the participant had pressed any of the 7 incorrect keys.

Experiment structure. Participants first performed 40 training trials: a short orientation only block ('O') of 10 trials, a short color only block of 10 trials and a short switch block ('S') of 20 trials. Afterwards, they performed 8 blocks of about 100 trials each in the order 'O-C-S-S-S-S-C-O' or 'C-O-S-S-S-S-O-C', with 30 seconds breaks in between blocks. Participants completed in total 800 non-aborted trials, approximately 200 in each one of the four conditions, Ori, Col, OriS and ColS (from 'S' blocks).

Target stimulus generation

The orientation and color of the target stimulus were based on the participant's previous responses according to an adaptive procedure, applied separately for each of the four conditions. We used Acerbi's Matlab implementation of the Ψ method called Psybayes [31], based on [81] with extentions to include the lapse rate [82]. This procedure maintains a posterior distribution over the parameters and updates it after each trial on which the participant pressed one of the 2 task-relevant buttons. The next stimulus value is chosen to minimize the entropy of the updated posterior given the stimulus, averaged over the participant's possible responses weighted by their respective probabilities [81]. Each generated on every trial a unitless value w within the range [-0.5, 0.5] that was converted to stimulus values: to $180\frac{w}{\pi}$ degrees in orientation trials and to $[L, (w+0.5)a_{\text{yellow}} + (0.5-w)a_{\text{blue}}, (w+0.5)b_{\text{yellow}} + (0.5-w)b_{\text{blue}}]$ in color trials. Thus, target stimulus values fell within an orientation range of -30 deg to 30 deg and within a blue to yellow range from [78 -30 -40] to [78 0 80]. For μ, we used a linear grid of 51 points from -0.5 to 0.5, for σ, a logarithmic grid of 25 points from 0.002 to 0.8 and for λ a linear grid of 25 points from 0 to 0.3.

Statistical analyses

For most metrics, we report median values and 95% bootstrapped confidence intervals. Across 50000 iterations, we took samples with replacement from and of the same size as the original data with Matlab's `randsample` and calculated the median of each of those sets of samples. The the 2.5th and 97.5th quantiles of the distribution of medians across iterations were taken as the 95% confidence intervals.

Three-way mixed-design ANOVA. To determine the differences between groups and the 2 experimental conditions of load and feature, we used three-way mixed-design ANOVA with two repeated measures, since we have one "between - participants" variable (group) and two "within - participants" factors (feature - Ori vs Col and load- No-switch vs Switch). Beforehand, we log transformed the measures that were lower bounded by 0.

When we assumed shared parameters between No-switch and Switch and thus we had only one "within - participants" factor, we used two-way mixed-design ANOVA. We implemented the ANOVAs in SPSS with "General linear model: repeated measures".

For post-hoc comparisons, we adjust the significance level according to the Sidak correction to $\alpha_{\text{sid}} = 1-(1-\alpha)^{\frac{1}{\text{number of comparisons}}}$. For the three-way mixed-design ANOVA, we performed, unless otherwise specified, 12 planned pairwise comparisons in Matlab: Wilcoxon rank-sum tests between groups (one for each condition, 4 in total), and Wilcoxon signed-rank tests for conditions within a group (4 per group, 8 in total). We used the Sidak correction for multiple comparisons, decreasing the significance level to $\alpha = 0.0043$ for post hoc comparisons following the three-way mixed-design ANOVA or respectively $\alpha = 0.0127$ following the two-way mixed-design ANOVA.

Pairwise correlations. To correct for multiple comparisons when examining the pairwise correlation matrix of the performance measures, we used a method from Nyholt et al. [83]. If M is the total number of measures, the number of effective comparisons will be decreased more if the measures are more highly correlated, as captured in a higher variance of the eigenvalues λ_{obs} of the correlation matrix, which we calculated with Matlab's function `eig`. Then, $M_{\text{eff}} = 1 + (M-1)\left(1 - \frac{\text{var}(\lambda_{\text{obs}})}{M}\right)$. As in [83], M_{eff} is used in the Sidak correction (a slightly less conservative alternative to Bonferroni), modifying the significance level to $\alpha_{\text{sid}} = 1-(1-\alpha)^{\frac{1}{M_{\text{eff}}}}$.

Linear regression. We implemented multivariate linear regression with Matlab's `fitlm`.

Logistic regression for classification. We fit the logistic regression coefficients with Matlab's `glmfit` with input 'binomial' and the link parameter 'logit'. For a given participant, we used the task metrics and the fitted coefficients with `glmval` to get p(Diagnosis), which was then thresholded at 0.5 to predict the 0 or 1 ADHD diagnosis.

Stratified 10-fold cross-validation. In order to assess the use of this logistic regression classifier for out-of-sample prediction, we calculated the cross-validated accuracy. We did stratified 10-fold cross-validation, in which each fold had 4 participants, 2 ADHD and 2 Controls; thus, we trained the classifier to find the coefficients over 36 participants and tested over 4 and calculated the mean accuracy across folds. We did 1000 runs of this stratified 10-fold cross-validation to allow for different random assignments of participants into folds and took the mean accuracy over runs.

Parameter fitting

Psychometric curves and parameters. We fitted psychometric curves to trials on which a participant pressed one of the 2 relevant buttons. s denotes the normalized stimulus value on a given trial (ranging between [-0.5, 0.5]). The psychometric curve has the following form [84]:

$$p(r = 1|s; \mu, \sigma, \lambda) = \frac{1}{2} \cdot \lambda + (1-\lambda) \cdot \Phi(s; \mu, \sigma), \tag{1}$$

where $r = 1$ stands for a response "clockwise" (orientation) or for "more yellow" (color). The parameters are the point of subjective equality (or bias), μ, the inverse slope (or noise) parameter, σ - which both are inputs to the Gaussian cumulative density function (Φ) - and the lapse rate, λ. We had 4 conditions, Ori, OriS, Col and ColS and thus 4 psychometric curves.

Parameter estimation and model choice. We performed maximum-likelihood estimation of the psychometric curve parameters μ, σ, and λ. The likelihood of a parameter combination is the probability of the data given that parameter combination; we denote the log likelihood by LL. We assumed that trials are independent of each other and thus we summed the log likelihoods across all trials. We fitted orientation and color trials separately; thus the following log likelihoods apply to either set of trials. In the main model, we assumed that μ and λ are shared across both load conditions (No-switch and Switch), whereas σ might differ. These assumptions had both a practical and a principled motivation. Assuming that parameters are shared

between conditions reduced the number of parameters and made parameter estimates more reliable. Moreover, if μ reflects an overall bias and λ a generic lapsing process, we did not expect them to change with load. For a model without these assumptions, and a model comparison, see Supplementary section "Psychometric curve parameters". The log likelihood for trials in a given feature dimensions becomes

$$\begin{aligned} \mathrm{LL}(\mu, \lambda, \sigma_{\text{No-switch}}, \sigma_{\text{Switch}}) &= \log p(\text{data} \mid \mu, \lambda, \sigma_{\text{No-switch}}, \sigma_{\text{Switch}}) \\ &= \sum_{\text{No-switch trials } j} \log\ p(r_j | s_j; \mu, \lambda, \sigma_{\text{No-switch}}) + \sum_{\text{Switch trials } j} \log\ p(r_j | s_j; \mu, \lambda, \sigma_{\text{Switch}}), \end{aligned} \tag{2}$$

where s_j and r_j are the stimulus and the participant's response on the jth trial, respectively. To estimate the parameters, we searched on a grid with 201 values in each dimension: for μ linearly spaced from -0.2 to 0.2, for λ logarithmically spaced from 0.0001 to 0.3, and for each σ logarithmically spaced from 0.002 and 0.5.

Reaction times. For fitting ex-Gaussian distributions to reaction times, we used a custom made script modeled after Zandbelt (2014): exgauss [85].

Data and code availability. Clinical data is not available beyond diagnosis labels, but behavioral data and experiment and analysis code is available upon request.

References

[1] Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V. & Pennington, B. F. Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry* **57**, 1336–1346 (2005).

[2] Boonstra, A. M., Oosterlaan, J., Sergeant, J. A. & Buitelaar, J. K. Executive functioning in adult ADHD: a meta-analytic review. *Psychological Medicine* **35**, 1097–1108 (2005).

[3] Castellanos, F. X. & Tannock, R. Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Reviews Neuroscience* **3**, 617–628 (2002).

[4] Kofler, M. J. *et al.* Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review* **33**, 795–811 (2013).

[5] Barkley, R. A. Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin* **121**, 65–94 (1997).

[6] Casey, B. *et al.* Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* **36**, 374–383 (1997).

[7] Booth, J. R. *et al.* Larger deficits in brain networks for response inhibition than for visual selective attention in attention deficit hyperactivity disorder (ADHD). *Journal of Child Psychology and Psychiatry* **46**, 94–111 (2005).

[8] Cepeda, N. J., Cepeda, M. L. & Kramer, A. F. *Journal of Abnormal Child Psychology* **28**, 213–226 (2000).

[9] Homack, S. A meta-analysis of the sensitivity and specificity of the stroop color and word test with children. *Archives of Clinical Neuropsychology* **19**, 725–743 (2004).

[10] King, J. A., Colla, M., Brass, M., Heuser, I. & von Cramon, D. Y. Inefficient cognitive control in adult ADHD: evidence from trial-by-trial Stroop test and cued task switching performance. *Behavioral and Brain Functions* **3**, 1 (2007).

[11] Halleland, H. B., Haavik, J. & Lundervold, A. J. Set-shifting in adults with ADHD. *Journal of the International Neuropsychological Society* **18**, 728–737 (2012).

[12] Sergeant, J. A., Geurts, H. & Oosterlaan, J. How specific is a deficit of executive functioning for attention-deficit/hyperactivity disorder? *Behavioural Brain Research* **130**, 3–28 (2002).

[13] Sergeant, J. A. Modeling attention-deficit/hyperactivity disorder: A critical appraisal of the cognitive-energetic model. *Biological Psychiatry* **57**, 1248–1255 (2005).

[14] Karalunas, S. L. & Huang-Pollock, C. L. Integrating impairments in reaction time and executive function using a diffusion model framework. *Journal of Abnormal Child Psychology* **41**, 837–850 (2013).

[15] Salum, G. A. *et al.* Mechanisms underpinning inattention and hyperactivity: neurocognitive support for ADHD dimensionality. *Psychological Medicine* **44**, 3189–3201 (2014).

[16] Ma, I., van Duijvenvoorde, A. & Scheres, A. The interaction between reinforcement and inhibitory control in ADHD: A review and research guidelines. *Clinical Psychology Review* **44**, 94–111 (2016).

[17] Mason, D. J., Humphreys, G. W. & Kent, L. S. Exploring selective attention in ADHD: visual search through space and time. *Journal of Child Psychology and Psychiatry* **44**, 1158–1176 (2003).

[18] Mullane, J. C. & Klein, R. M. Literature review: Visual search by children with and without ADHD. *Journal of Attention Disorders* **12**, 44–53 (2008).

[19] Huang-Pollock, C. L. & Nigg, J. T. Searching for the attention deficit in attention deficit hyperactivity disorder: The case of visuospatial orienting. *Clinical Psychology Review* **23**, 801–830 (2003).

[20] Cubillo, A. *et al.* Reduced activation and inter-regional functional connectivity of fronto-striatal networks in adults with childhood attention-deficit hyperactivity disorder (ADHD) and persisting symptoms during tasks of motor inhibition and cognitive switching. *Journal of Psychiatric Research* **44**, 629–639 (2010).

[21] Rubia, K. *et al.* Disorder-specific inferior prefrontal hypofunction in boys with pure attention-deficit/hyperactivity disorder compared to boys with pure conduct disorder during cognitive flexibility. *Human Brain Mapping* **31**, 1823–1833 (2010).

[22] Roberts, M., Ashinoff, B., Castellanos, J. & Carrasco, M. Endogenous and exogenous covert attention are functionally intact in adults with adhd. *Vision Sciences Society* (2017).

[23] Micoulaud-Franchi, J.-A. *et al.* Perceptual abnormalities related to sensory gating deficit are core symptoms in adults with ADHD. *Psychiatry Research* **230**, 357–363 (2015).

[24] Bijlenga, D., Tjon-Ka-Jie, J., Schuijers, F. & Kooij, J. Atypical sensory profiles as core features of adult ADHD, irrespective of autistic symptoms. *European Psychiatry* **43**, 51–57 (2017).

[25] Fuermaier, A. B. M. *et al.* Perception in attention deficit hyperactivity disorder. *ADHD Attention Deficit and Hyperactivity Disorders* (2017).

[26] Kim, S. *et al.* Color vision in ADHD: Part 2 - does attention influence color perception? *Behavioral and Brain Functions* **10**, 39 (2014).

[27] Friedman-Hill, S. R. *et al.* What does distractibility in ADHD reveal about mechanisms for top-down attentional control? *Cognition* **115**, 93–103 (2010).

[28] Stevens, A. A. *et al.* Increased sensitivity to perceptual interference in adults with attention deficit hyperactivity disorder. *Journal of the International Neuropsychological Society* (2012).

[29] McAvinue, L. P. *et al.* A componential analysis of visual attention in children with ADHD. *Journal of Attention Disorders* **19**, 882–894 (2012).

[30] Kim, S. *et al.* Colour vision in ADHD: Part 1 - testing the retinal dopaminergic hypothesis. *Behavioral and Brain Functions* **10**, 38 (2014).

[31] Acerbi, L. Bayesian adaptive stimulus placement of psychometric function for matlab (2016). URL `https://github.com/lacerbi/psybayes`.

[32] Wiecki, T. V., Poland, J. & Frank, M. J. Model-based cognitive neuroscience approaches to computational psychiatry. *Clinical Psychological Science* **3**, 378–399 (2015).

[33] Castellanos, F. X., Sonuga-Barke, E. J., Milham, M. P. & Tannock, R. Characterizing cognition in ADHD: beyond executive dysfunction. *Trends in Cognitive Sciences* **10**, 117–123 (2006).

[34] Karalunas, S. L., Huang-Pollock, C. L. & Nigg, J. T. Decomposing attention-deficit/hyperactivity disorder (ADHD)-related effects in response speed and variability. *Neuropsychology* **26**, 684–694 (2012).

[35] Leth-Steensen, C., Elbaz, Z. K. & Douglas, V. I. Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychologica* **104**, 167–190 (2000).

[36] Douglas, V. I. Cognitive control processes in attention deficit/hyperactivity disorder. In *Handbook of Disruptive Behavior Disorders*, 105–138 (Springer US, 1999).

[37] Kingdom, P.-N., F.A. *Psychophysics: A Practical Introduction* (Academic Press, London, 2009).

[38] Roth, R. M., Lance, C. E., Isquith, P. K., Fischer, A. S. & Giancola, P. R. Confirmatory factor analysis of the behavior rating inventory of executive function-adult version in healthy adults and application to attention-deficit/hyperactivity disorder. *Archives of Clinical Neuropsychology* **28**, 425–434 (2013).

[39] Adler, L. & Cohen, J. Diagnosis and evaluation of adults with attention-deficit/hyperactivity disorder. *Psychiatric Clinics of North America* **27**, 187–201 (2004).

[40] Adler, L. A. *et al.* The structure of adult ADHD. *International Journal of Methods in Psychiatric Research* **26**, e1555 (2017).

[41] Liu, T. Cortical mechanisms of feature-based attentional control. *Cerebral Cortex* **13**, 1334–1343 (2003).

[42] Greenberg, A. S., Esterman, M., Wilson, D., Serences, J. T. & Yantis, S. Control of spatial and feature-based attention in frontoparietal cortex. *Journal of Neuroscience* **30**, 14330–14339 (2010).

[43] Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).

[44] Siegel, M., Buschman, T. J. & Miller, E. K. Cortical information flow during flexible sensorimotor decisions. *Science* **348**, 1352–1355 (2015).

[45] Haigh, S. M. *et al.* No difference in cross-modal attention or sensory discrimination thresholds in autism and matched controls. *Vision Research* **121**, 85–94 (2016).

[46] Kim, S., Chen, S. & Tannock, R. Visual function and color vision in adults with attention-deficit/hyperactivity disorder. *Journal of Optometry* **7**, 22–36 (2014).

[47] Huang-Pollock, C. *et al.* Using the diffusion model to explain cognitive deficits in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology* **45**, 57–68 (2016).

[48] Ziegler, S., Pedersen, M. L., Mowinckel, A. M. & Biele, G. Modelling ADHD: A review of ADHD theories through their predictions for computational models of decision-making and reinforcement learning. *Neuroscience & Biobehavioral Reviews* **71**, 633–656 (2016).

[49] Hauser, T. U. *et al.* Role of the medial prefrontal cortex in impaired decision making in juvenile attention-deficit/hyperactivity disorder. *JAMA Psychiatry* **71**, 1165 (2014).

[50] Hauser, T. U., Fiore, V. G., Moutoussis, M. & Dolan, R. J. Computational psychiatry of ADHD: Neural gain impairments across marrian levels of analysis. *Trends in Neurosciences* **39**, 63–73 (2016).

[51] Drugowitsch, J., Wyart, V., Devauchelle, A.-D. & Koechlin, E. Computational precision of mental inference as critical source of human choice suboptimality. *Neuron* **92**, 1398–1411 (2016).

[52] Thomas, M. L. *et al.* Modeling deficits from early auditory information processing to psychosocial functioning in schizophrenia. *JAMA Psychiatry* **74**, 37 (2017).

[53] Frank, M. J., Santamaria, A., OReilly, R. C. & Willcutt, E. Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology* **32**, 1583–1599 (2006).

[54] del Campo, N., Chamberlain, S. R., Sahakian, B. J. & Robbins, T. W. The roles of dopamine and noradrenaline in the pathophysiology and treatment of attention-deficit/hyperactivity disorder. *Biological Psychiatry* **69**, e145–e157 (2011).

[55] Wimmer, R. D. *et al.* Thalamic control of sensory selection in divided attention. *Nature* **526**, 705–709 (2015).

[56] Schmitt, L. I. & Halassa, M. M. Interrogating the mouse thalamus to correct human neurodevelopmental disorders. *Molecular Psychiatry* **22**, 183–191 (2016).

[57] Young, A. & Wimmer, R. D. Implications for the thalamic reticular nucleus in impaired attention and sleep in schizophrenia. *Schizophrenia Research* **180**, 44–47 (2017).

[58] Ivanov, I. *et al.* Morphological abnormalities of the thalamus in youths with attention deficit hyperactivity disorder. *American Journal of Psychiatry* **167**, 397–408 (2010).

[59] Batty, M. J. *et al.* Morphological abnormalities in prefrontal surface area and thalamic volume in attention deficit/hyperactivity disorder. *Psychiatry Research: Neuroimaging* **233**, 225–232 (2015).

[60] Bailey, T. & Joyce, A. The Role of the Thalamus in ADHD Symptomatology and Treatment. *Applied Neuropsychology: Child* (2015).

[61] Leventhal, A. G., Schmolesky, M. T., Wang, Y. & Pu, M. Degradation of stimulus selectivity of visual cortical cells in senescent rhesus monkeys. *Nature Neuroscience* **3**, 384–390 (2000).

[62] Wells, M. F., Wimmer, R. D., Schmitt, L. I., Feng, G. & Halassa, M. M. Thalamic reticular impairment underlies attention deficit in ptchd1y/- mice. *Nature* **532**, 58–63 (2016).

[63] Leo, D. & Gainetdinov, R. R. Transgenic mouse models for ADHD. *Cell and Tissue Research* **354**, 259–271 (2013).

[64] Mueller, A., Hong, D. S., Shepard, S. & Moore, T. Linking ADHD to the neural circuitry of attention. *Trends in Cognitive Sciences* (2017).

[65] Thome, J. *et al.* Biomarkers for attention-deficit/hyperactivity disorder (adhd). a consensus report of the wfsbp task force on biological markers and the world federation of adhd. *The World Journal of Biological Psychiatry* **13**, 379–400 (2012).

[66] Ogundele, M. O., Ayyash, H. F. & Banerjee, S. Role of computerised continuous performance task tests in ADHD. *Progress in Neurology and Psychiatry* **15**, 8–13 (2011).

[67] Fried, M. *et al.* ADHD subjects fail to suppress eye blinks and microsaccades while anticipating visual stimuli but recover with medication. *Vision research* **101**, 62–72 (2014).

[68] Panagiotidi, M., Overton, P. & Stafford, T. Increased microsaccade rate in individuals with adhd traits. *Journal of Eye Movement Research* (2017).

[69] Wainstein, G. *et al.* Pupil size tracks attentional performance in attention-deficit/hyperactivity disorder. *Scientific Reports* **7** (2017).

[70] Mous, S. E. *et al.* Cortical thickness and inattention/hyperactivity symptoms in young children: a population-based study. *Psychological Medicine* **44**, 3203–3213 (2014).

[71] Montague, P. R., Dolan, R. J., Friston, K. J. & Dayan, P. Computational psychiatry. *Trends in Cognitive Sciences* **16**, 72–80 (2012).

[72] Redish, A. D. & Gordon, J. A. *Computational Psychiatry: New Perspectives on Mental Illness* (MIT Press, 2016).

[73] Vinogradov, S. The golden age of computational psychiatry is within sight. *Nature Human Behaviour* **1**, 0047 (2017).

[74] Kessler, R. C. *et al.* The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychological Medicine* **35**, 245–256 (2005).

[75] Kessler, R. C. *et al.* The prevalence and correlates of adult ADHD in the united states: Results from the national comorbidity survey replication. *American Journal of Psychiatry* **163**, 716–723 (2006).

[76] Brainard, D. H. The psychophysics toolbox. *Spatial Vision* **10**, 433–436 (1997).

[77] Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* **10**, 437–442 (1997).

[78] Kleiner, M. *et al.* Whats new in psychtoolbox-3. *Perception* **36**, 1 (2007).

[79] Cornelissen, F. W., Peters, E. M. & Palmer, J. The eyelink toolbox: Eye tracking with MATLAB and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers* **34**, 613–617 (2002).

[80] Liu, T., Stevens, S. T. & Carrasco, M. Comparing the time course and efficacy of spatial and feature-based attention. *Vision research* **47**, 108–113 (2007).

[81] Kontsevich, L. L. & Tyler, C. W. Bayesian adaptive estimation of psychometric slope and threshold. *Vision research* **39**, 2729–2737 (1999).

[82] Prins, N. The adaptive psi method and the lapse rate. *Journal of Vision* **12**, 322–322 (2012).

[83] Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics* **74**, 765–769 (2004).

[84] Wichmann, F. A. & Hill, N. J. The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & Psychophysics* **63**, 1293–1313 (2001).

[85] Zandbelt, B. exgauss: a matlab toolbox for fitting the ex-gaussian distribution to response time data (2014).

[86] Hurvich, C. M. & Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika* **76**, 297–307 (1989).

[87] Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464 (1978).

[88] Tannock, R. & Banaschewski, T. Color naming deficits and attention-deficit/hyperactivity disorder: a retinal dopaminergic hypothesis. *Behavioral and Brain Functions* (2006).

Supplementary material

S1 Demographic and clinical information

	Demographics		
Group	Gender	Age	# White
Control	$11F, 9M$	32.5 ± 6.1	11
ADHD	$12F, 8M$	35.3 ± 10.0	9

Table S1: **Demographic information of participants.**
Values represent mean and standard deviation.

	ADHD scale scores					
Group	ACDS	ACDS2	ASRS	AISRS	MCI	GEC
Control	25.1 ± 4.1	1.3 ± 1.7	19.3 ± 8.8	6.9 ± 4.7	45.7 ± 8.1	45.6 ± 8.5
ADHD	52.8 ± 6.8	14.4 ± 2.5	49.7 ± 6.6	36.5 ± 7.9	73.7 ± 9.0	71.2 ± 6.8
Wilcoxon rank-sum p values	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-5}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$

Table S2: **Psychiatric scores of participants.**
Values represent mean and standard deviation. ACDS denotes ACDS B1-B18, and ACDS2 B22-B39.

	ASRS	AISRS	ACDS	ACDS2	GEC
ASRS					
AISRS	0.85***				
ACDS	0.90***	0.96***			
ACDS2	0.85***	0.96***	0.94***		
GEC	0.85***	0.77***	0.82***	0.81***	
MCI	0.81***	0.81***	0.84***	0.83***	0.95***

Table S3: **Spearman correlations across the scores for all diagnosis scales.**
10 participants were excluded from this table because not all records were available. However, ACDS, ACDS2, GEC and MCI were available for all participants.

S2 Further characterization of responses

Accuracy was maintained approximately constant across participants and conditions (mean ± sem: 0.811 ± 0.005) due to the Psybayes method of adaptive thresholding (Figure S1a). We further characterized the TIMO responses, first with a breakdown by error type (Figure S1b) and then according to the type of the previous trial (Figure S1c). In the Switch trials, the majority of TIMO seem to be feature errors, possibly because the mapping from left/right visual field to left/right keyboard was more intuitive than from feature dimension to top/down of a keyboard (Figure S1b). No clear pattern emerged from breaking down by the type of the previous trial (Figure S1c).

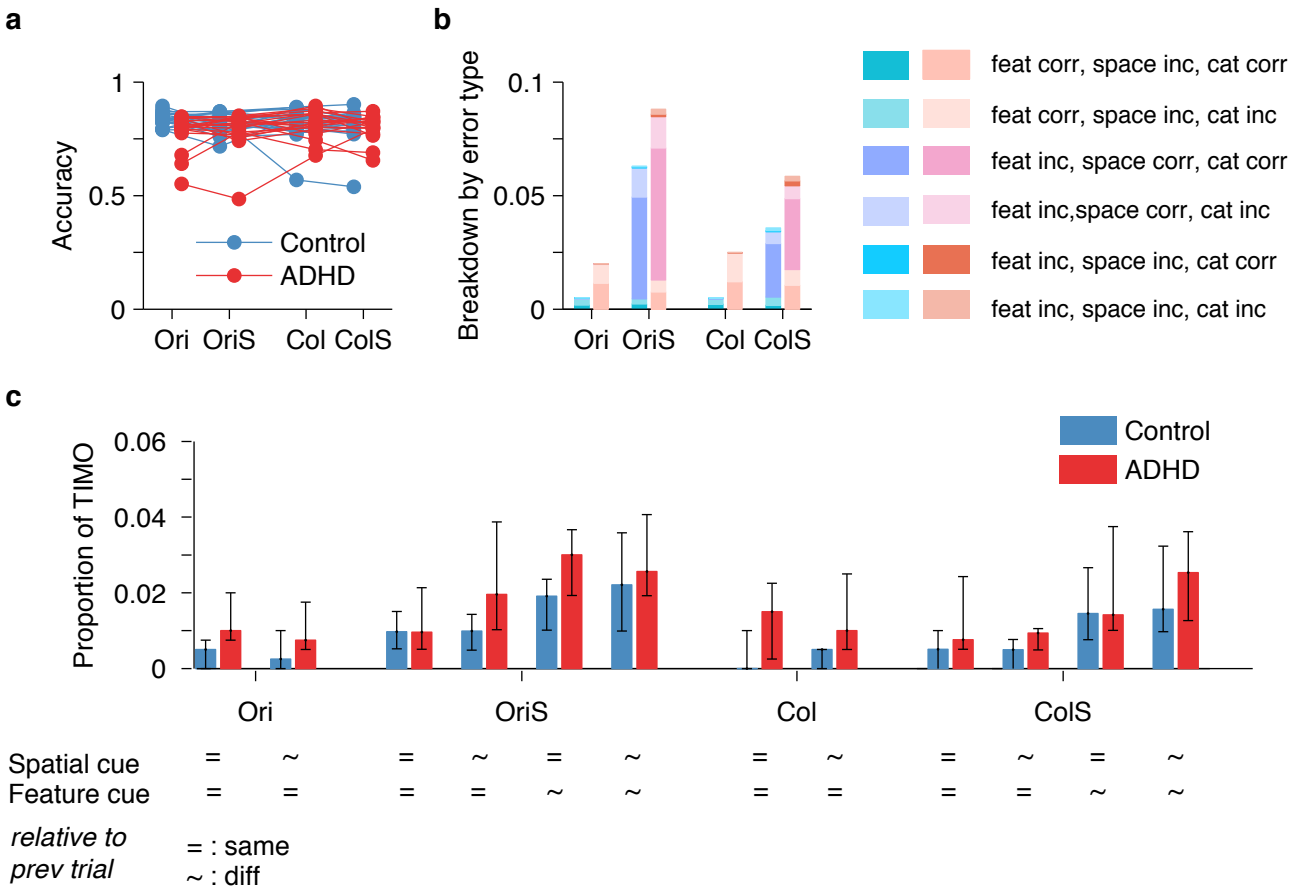

Figure S1: **Further characterization of responses.** **(a)** Accuracy as proportion correct on the trials when participants selected one of the 2 relevant keys. **(b)** The task irrelevant motor output from Figure 1, broken down by error type. For every condition, the first bar is Controls, and the second one ADHD. **(c)** Same, broken down by the type of the previous trial.

S3 Further information on reaction times

S3.1 ex-Gaussian model

ex-Gaussian distributions are commonly fitted to reaction time data and are defined by adding 2 random variables, a Gaussian with parameters μ and σ and an exponential with parameter τ. While in our data τ showed an effect of group, neither log μ_{RT} ($F(1,38) = 0.05, p = 0.83, \eta_p^2 = 0.001$) nor log σ_{RT} ($F(1,38) = 0.27, p = 0.61, \eta_p^2 = 0.007$) did, consistent with other studies that showed significant effects of group on τ but not on μ_{RT} (see [4] for a meta-analysis).

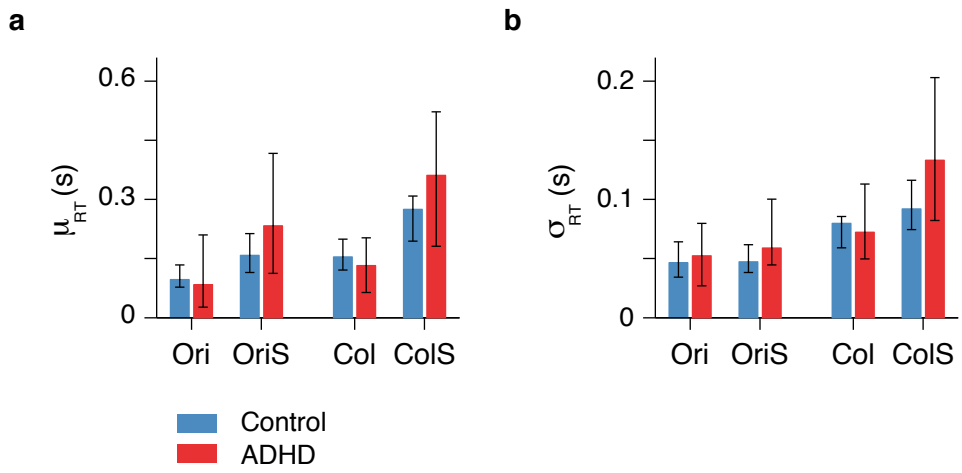

Figure S2: **ex-Gaussian parameters fitted to the reaction time distributions across conditions and groups.** **(a)** Gaussian mean μ_{RT}, **(b)** Gaussian standard deviation σ_{RT}, both for each task condition.

S3.2 Alternative models

While ex-Gaussian distributions are routinely used to fit reaction times, they are rarely compared to alternative distributions. We used the corrected Akaike Information Criterion (AICc) and the Bayesian Information criterion (BIC) to compare the ex-Gaussian fits with the fits of 2 other distributions on the positive real line: log-Normal and Gamma. These metrics are defined as $\mathrm{AICc} = -2\mathrm{LL}^* + 2k + \frac{2k(k+1)}{n_{\mathrm{trials}}-k-1}$ [86] and $\mathrm{BIC} = -2\mathrm{LL}^* + k \log n_{\mathrm{trials}}$ [87], respectively, where LL^* is the maximum log likelihood, k is the number of free parameters, and n_{trials} is the number of trials. We found that indeed the ex-Gaussian distribution fitted better than both the log-Normal (in median by 611 according to AICc and by 607 according to BIC) and the Gamma distribution (in median by 50 according to AICc and by 45 according to BIC); see Figure S3 for individual subjects.

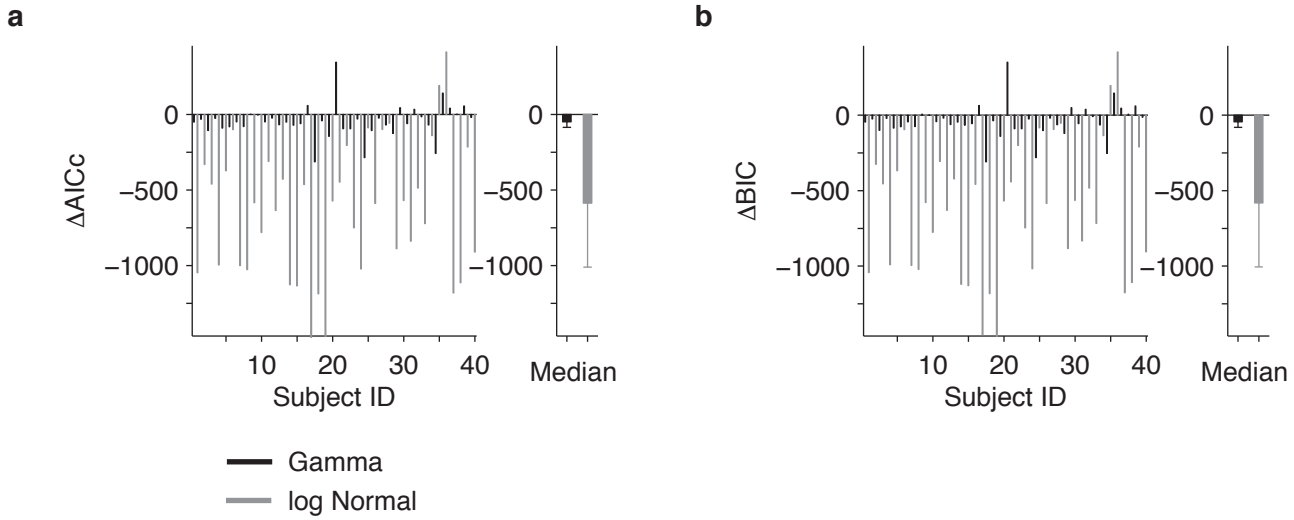

Figure S3: **Model comparison justifies the parametrization of reaction times with the ex-Gaussian distribution.** **(a)** The ex-Gaussian model has the lowest AICc across the population **(Right)** and for almost all individual subjects **(Left)**. **(b)** Same result for BIC.

S3.3 Non-parametric measure of RT variability

We complemented the results about RT τ (Figure 2) with a non-parametric robust measure of intra-individual reaction time variability, the reaction time inter-quartile range (iqr) (Figure S4). Three-way mixed-design ANOVA on log RT iqr's revealed a significant effect of group ($F(1,38) = 5.13, p = 0.029, \eta_p^2 = 0.12$), load ($F(1,38) = 18.84, p < 0.001, \eta_p^2 = 0.33$), feature ($F(1,38) = 21.38, p < 0.001, \eta_p^2 = 0.36$), and a significant load $\times$ feature interaction ($F(1,38) = 22.12, p < 0.001, \eta_p^2 = 0.37$). No other two-way interaction nor the three-way interaction were significant ($p > 0.36$). After Sidak correction ($\alpha = 0.0043$), none of the between-groups comparisons were significant. Within Controls, the effects of load and feature on log RT iqr were significant for Ori vs OriS ($p = 0.0015$) and Ori vs Col ($p < 0.001$); within ADHD, the only significant effect was of load for Ori vs OriS ($p = 0.002$).

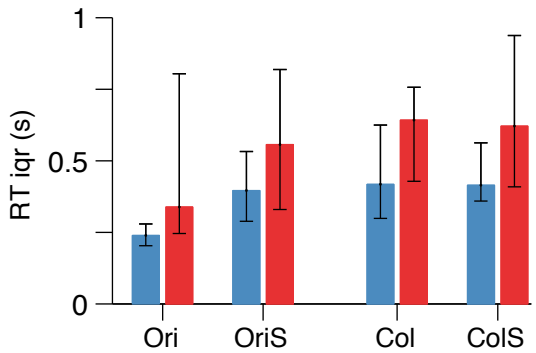

Figure S4: **Reaction time variability is higher in ADHD also according to a non-parametric metric, RT iqr.**

S4 Psychometric curve parameters

S4.1 PSE

Figure S6d shows the estimates of μ (PSE) in the "shared" (main) model. Two-way mixed-design ANOVA on μ with within-group factor feature showed an effect of group ($F(1,38) = 9.47, p = 0.004, \eta_p^2 = 0.2$), but no significant effect of feature ($F(1,38) = 1.17, p = 0.28, \eta_p^2 = 0.03$) and not a significant interaction. After Sidak correction ($\alpha = 0.0253$), no effects were significant. We chose to interpolate color values between blue and yellow since the S-cone pathway is of special interest in ADHD [88]. While we found an overall group effect on μ, after Sidak correction the post hoc effect for color failed to reach significance, thus making our results at this point inconclusive about whether ADHD participants have different S-cone dependent color processing.

S4.2 "Full" model, 12 parameters

While in the main or "shared" model (Figure 3) we assumed that μ and λ were shared within a feature across load conditions, in the "full" model we did not constrain any parameters, yielding 12 parameters total.

As expected, the "full" model captured the data at least as well as the "shared" model. However, the "shared" model fitted the data slightly better (in median by -1.5 according to AICc) or better (in median by -20 according to BIC) than the "full" model (Figure S5). This confirmed the plausibility of the shared-parameters assumption in the main model.

We also performed three-way mixed-design ANOVA on the parameter estimates from the "full" model (Figure S6B). Just like in the "shared" model, we found a significant effect of group for log perceptual variability (σ) ($F(1,38) = 5.21, p = 0.028, \eta_p^2 = 0.12$), a significant effect of feature ($F(1,38) = 37.11, p < 0.0001, \eta_p^2 = 0.49$), but no significant effect of load ($F(1,38) = 0.03, p = 0.87, \eta_p^2 = 0.001$). Neither of the two-way interactions nor the three-way interaction were significant ($p > 0.06$). After Sidak correction ($\alpha = 0.0065$, 8 comparisons, since we excluded across feature comparisons, as in the main model) we found a between-group effect for Ori

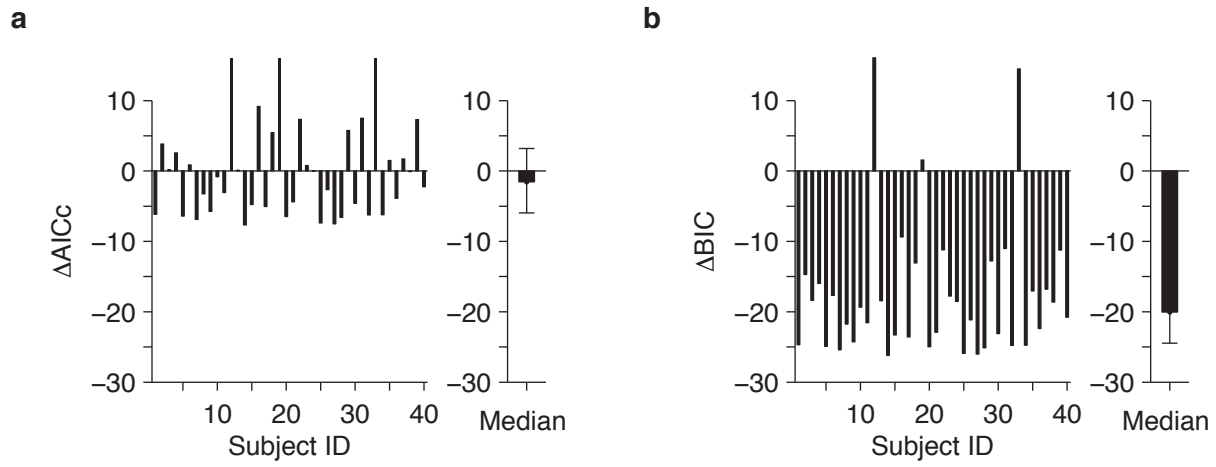

Figure S5: **Model comparison justifies using the "shared" model. (a)** AICc of the "shared" model minus AICc of the "full" model for **(Left)** individual subjects and **(Right)** Group - median and 95% bootstrapped confidence intervals. **(b)** Same for BIC.

with $p < 0.0001$ and OriS ($p < 0.0025$), but no significant effects of group for neither Col ($p = 0.0063$) nor ColS ($p = 0.47$) (Figure S6).

For the log lapse λ, we found a significant effect of feature ($F(1, 38) = 25.88, p < 0.0001, \eta_p^2 = 0.40$) and a significant feature $\times$ group interaction ($F(1, 38) = 6.01, p = 0.02, \eta_p^2 = 0.14$); nothing else was significant ($p > 0.09$). After Sidak correction ($\alpha = 0.0043$, all 12 comparisons make sense since λ is unitless), no between-group comparisons were significant ($p > 0.02$). Within Controls, the feature comparisons Ori vs Col and OriS vs ColS were significant ($p < 0.001$), but not the load ones. Within ADHD, neither the feature nor the load comparisons reached significance ($p > 0.02$).

For the PSE μ, like in the "shared" model, we found a significant effect of group ($F(1, 38) = 10.85, p = 0.002, \eta_p^2 = 0.22$) and also a significant group $\times$ load $\times$ feature interaction ($F(1, 38) = 8.42, p = 0.006, \eta_p^2 = 0.18$), nothing else reaching significance ($p > 0.09$). After Sidak correction ($\alpha = 0.0065$, as for σ), we found a significant difference between ADHD and Control for ColS ($p = 0.002$), but not for Col ($p = 0.18$) and not for Ori or OriS ($p > 0.12$). Again, these results cannot provide robust support for ADHD participants having different S-cone dependent color processing.

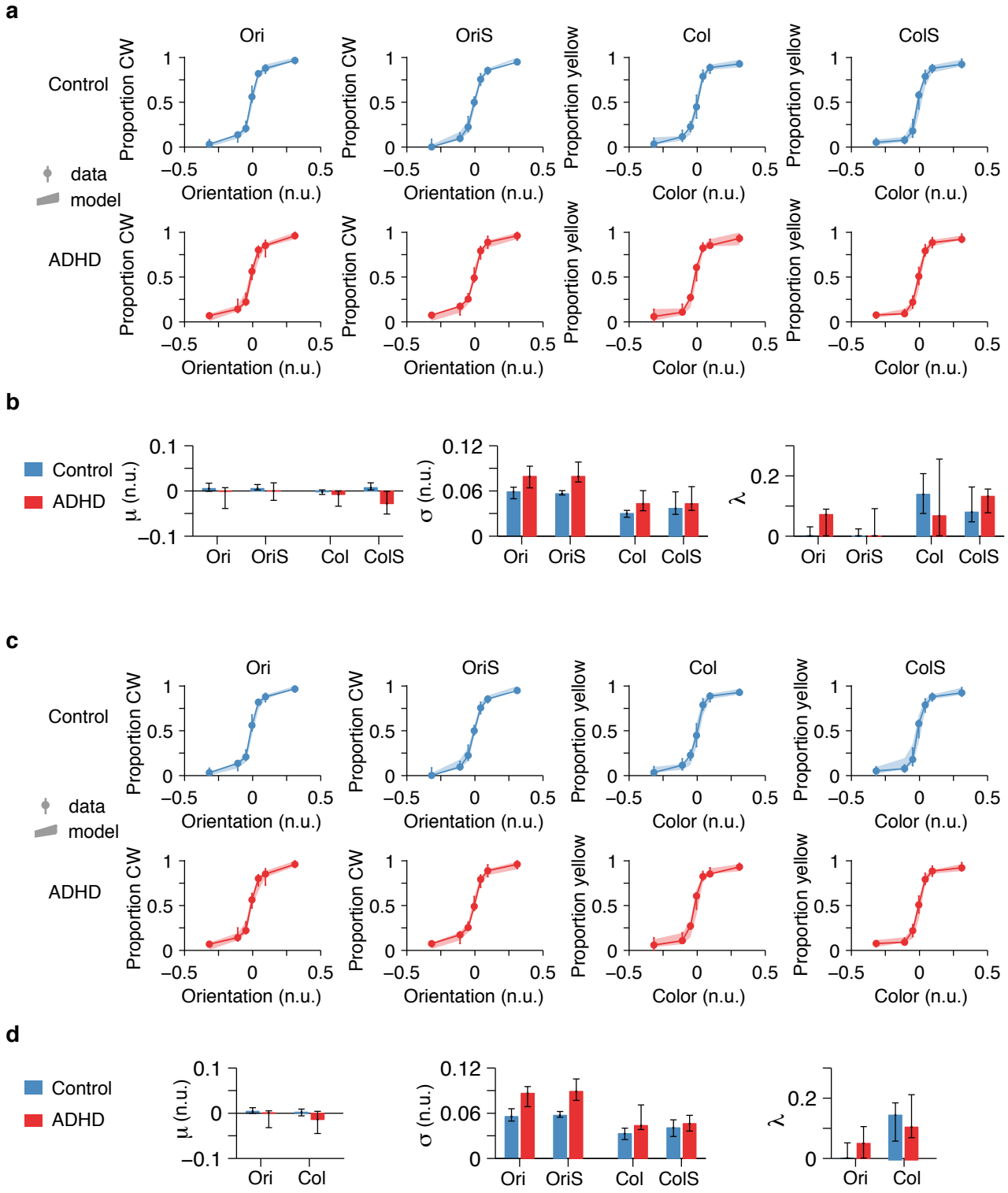

Figure S6: **(a)** "Full model" (12 parameters total): data and fitted psychometric curves. Solid circles with error bars show median and 95% bootstrapped confidence intervals, while shaded area show the same for model predictions. The data was binned into 7 quantiles. Since the Bayesian adaptive method presented each participant in each condition with an unique set of stimuli, the midpoint stimulus values of the quantile bins differed for each. However, for ease of visualization, here we place the midpoints stimulus values for each bin as the midpoints obtained from binning into 7 quantiles the entire stimulus set concatenated across participants and conditions. **(b)** "Full model": MLE parameter fits, **(c)** "Shared model" (8 parameters total): data and fitted psychometric curves. **(d)** "Shared model": MLE parameter fits.

S5 Effect of eye tracking

A possible concern is that half of the participants in each group were eye tracked, while half were not. If an eye-tracked participant broke fixation, they had to redo the trial. As a result, the eye-tracked participants started more trials (mean and SD for eye-tracked: 1047 ± 201 trials; non-eye-tracked always completed 800 trials). Thus, a concern could be that differences in task metrics could simply arise due to the experiment being longer and as a result more tiresome. We examined each of the average task metrics within each group, separately for the eye-tracked participants and the non-eye-tracked ones and found no significant differences (Wilcoxon rank-sum $p > 0.13$) (Figure S7).

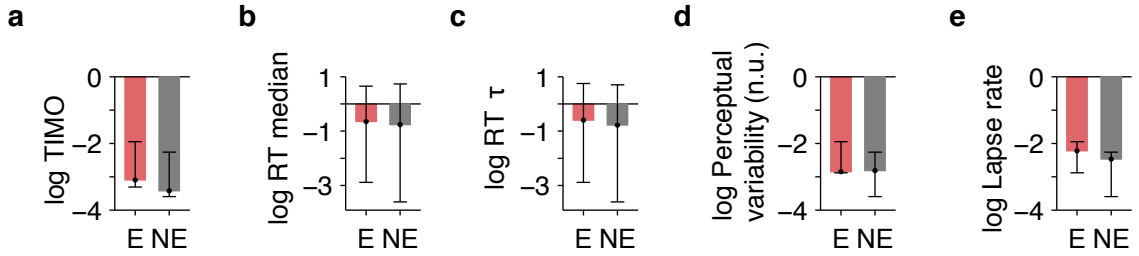

Figure S7: **No significant difference between eye tracked (E) and non-eye tracked (NE) participants on behavioral task metrics: (a) TIMO, (b) RT median, (c) RT τ, (d) Perceptual variability, (e) Lapse rate.** Bars represent medians and error bars bootstrapped 95% confidence intervals.

S6 Breakdown of correlations

S6.1 By group

In Figure S8, we show the points that make up the correlations from Table 1, color coded by group. In Table S4 we show the pairwise correlations across task metrics separately within the Control group and within the ADHD group. Here, the only group specific correlations that survive the multiple-comparisons correction are RT with RT τ and GEC with ACDS.

In addition, we attempted to determine whether for a given pair of task metrics, their correlation within the ADHD group is different from their correlation within the Control group. To do this, we compared the difference between the actual correlations to a distribution of differences between correlations obtained by shuffling the ADHD and Control labels. We did not find significant differences between the ADHD and Control correlations for any pair of task metrics ($p > 0.04$).

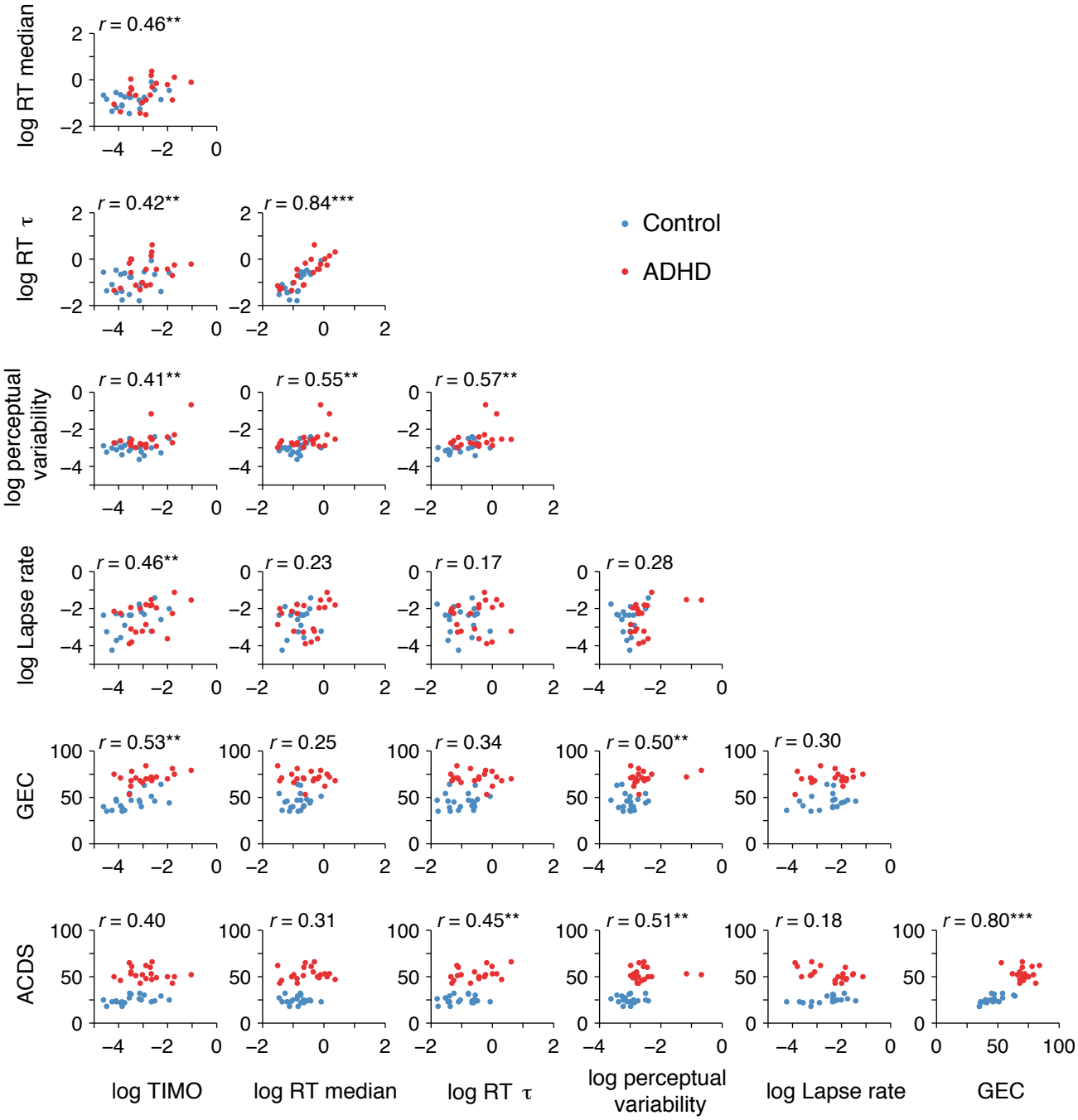

Figure S8: **Dots: pairwise task metrics, color coded by group**. We also show here the correlations collapsed across groups, as in Table 1. * depicts $0.001 < p < 0.0089$ (since $\alpha_{\text{Sidak}} = 0.0089$ after multiple-comparisons correction) and ** depicts $p < 0.001$.

Table S4: No evident pattern of group specific correlations. Pairwise Spearman correlations across task metrics (both behavioral and clinical), as in Table 1, but divided by group. Boldfaced is significant after multiple-comparisons correction, $\alpha = 0.0083$ for Control and $\alpha = 0.0082$ for ADHD. (see Methods)

(a) **Control**

	TIMO	RT	RT τ	Perceptual variability (σ)	Lapse (λ)	GEC
TIMO						
RT	$r = 0.25$ $p = 0.27$					
RT τ	$r = 0.13$ $p = 0.59$	$\boldsymbol{r}$**=0.81** $p < 0.0001$				
Perceptual variability (σ)	$r = -0.02$ $p = 0.92$	$r = 0.52$ $p = 0.02$	$r = 0.50$ $p = 0.03$			
Lapse rate (λ)	$r = 0.49$ $p = 0.03$	$r = 0.22$ $p = 0.34$	$r = 0.19$ $p = 0.42$	$r = -0.05$ $p = 0.84$		
GEC	$r = 0.52$ $p = 0.02$	$r = 0.16$ $p = 0.49$	$r = 0.15$ $p = 0.52$	$r = -0.12$ $p = 0.61$	$r = 0.25$ $p = 0.29$	
ACDS	$r = 0.49,$ $p = 0.03$	$r = -0.09$ $p = 0.67$	$r = 0.03$ $p = 0.89$	$r = -0.09$ $p = 0.69$	$r = 0.46$ $p = 0.04$	$\boldsymbol{r}$**=0.70** $p < 0.0001$

(b) **ADHD**

	TIMO	RT	RT τ	Perceptual variability (σ)	Lapse (λ)	GEC
TIMO						
RT	$r = 0.48$ $p = 0.03$					
RT τ	$r = 0.36$ $p = 0.12$	$\boldsymbol{r}$**=0.82** $p < 0.0001$				
Perceptual variability (σ)	$r = 0.46$ $p = 0.04$	$r = 0.52$ $p = 0.02$	$r = 0.46$ $p = 0.04$			
Lapse rate (λ)	$r = 0.40$ $p = 0.08$	$r = 0.38$ $p = 0.10$	$r = 0.12$ $p = 0.61$	$r = 0.32$ $p = 0.16$		
GEC	$r = 0.37$ $p = 0.11$	$r = -0.10$ $p = 0.68$	$r = 0.19$ $p = 0.42$	$r = 0.15$ $p = 0.54$	$r = 0.18$ $p = 0.44$	
ACDS	$r = -0.18$ $p = 0.45$	$r = 0.14$ $p = 0.56$	$r = 0.36$ $p = 0.11$	$r = 0.03$ $p = 0.87$	$r = -0.35$ $p = 0.12$	$r = -0.09$ $p = 0.07$

S6.2 By condition

In Table 1, for each participant, we averaged each behavioral metric across all four conditions. In Figure S9, we present the correlations of perceptual variability with TIMO, RT and RT τ broken down by condition. The correlations that survived multiple-comparisons correction are between perceptual variability (σ) and TIMO in the Ori condition, between σ and RT in the OriS and σ and RT τ in Ori and Col. Overall, we cannot conclude much from these patterns of results.

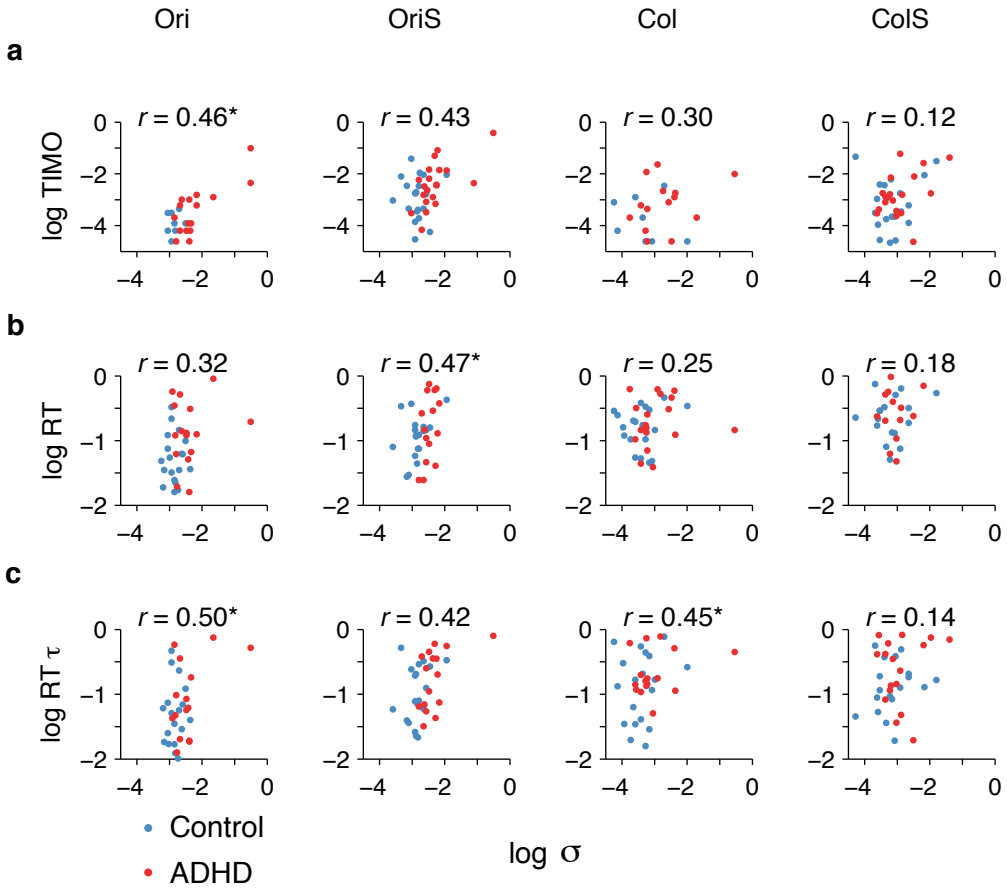

Figure S9: **Spearman correlations broken down by conditions show no conclusive pattern. We show the correlations of log perceptual variability with (a) log TIMO, (b) log RT median and (c) log RT τ.** *'s represent correlations with p values smaller than $\alpha_{\text{Sidak}} = 0.0043$.

S7 Prediction of clinical variables

S7.1 Linear regression: prediction of clinical metrics GEC and ACDS from behavioral metrics

Table S5: **Linear regression coefficients, depicted as mean ± sem, for GEC and ACDS with task metrics**.

(a) **GEC ∼ log perceptual variability + log TIMO + log RT + log RT τ + log lapse rate.**

	coefficient	t value	p value
log intercept	103 ± 12	8.50	$< 10^{-9}$***
log TIMO	8.0 ± 3.2	2.53	0.016*
log RT median	-10.3 ± 7.6	-1.35	0.18
log RT τ	8.1 ± 6.1	1.32	0.19
log σ	6.1 ± 4.9	1.23	0.22
log λ	1.3 ± 1.8	0.75	0.46

(b) **ACDS ∼ log perceptual variability + log TIMO + log RT + log RT τ + log lapse rate.**

coefficient	t value	p value
73 ± 13	5.77	$< 10^{-5}$***
3.4 ± 3.3	1.04	0.31
-11.3 ± 8.0	-1.41	0.15
14.4 ± 6.4	2.26	0.031*
6.8 ± 5.1	1.33	0.19
0.5 ± 1.9	0.29	0.78

S7.2 Logistic regression: prediction of diagnosis from behavioral metrics

Figure S10a shows the placement of the following three logistic regression classifiers in the space of log perceptual variability (σ) and log TIMO. Classification accuracy was 77.5% with perceptual variability only, 62.5% with TIMO only and 82.5% with both. The addition of the other task variables, RT, RT τ and lapse, did not improve the accuracy further (80.0%). With 1000 runs of stratified 10-fold cross-validation, mean accuracy was 77.1% with perceptual variability only, 63.1% with TIMO only, 77.8% with both and 70.0% with all metrics.

In addition to thresholding at 0.5 to get diagnosis and subsequently accuracy as above, we also thresholded p(Diagnosis) at linearly spaced values between 0 and 1 and plotted the resulting receiver operating characteristic (ROC) curves, both without (Figure S10b) and with stratified 10-fold cross-validation (Figure S10c). As expected, the ROC curve for the classifier all metrics shows the best performance (highest area under the curve (AUC)) without cross-validation, but its performance degrades for out-of-sample predictions in the cross-validated case.

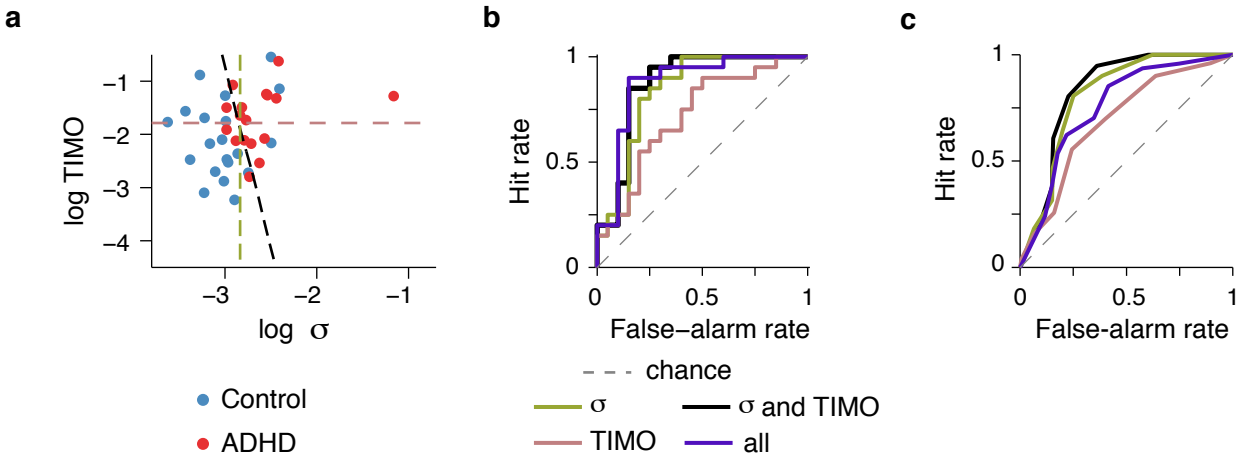

Figure S10: **Logistic regression for classification of participants into ADHD and Controls based on task metrics.** **(a)** Dots: combinations of log TIMO and log perceptual variability (σ) across participants. Dashed lines: logistic regression classifiers trained on log σ only (olive), TIMO only (old rose) and both (black). **(b)** Full ROC curves obtained by varying the diagnosis threshold for the three classifiers in **(a)**, as well as for one based on all 5 behavioral metrics (purple). **(c)** Full ROC curves, this time with stratified 10-fold cross-validation, for the same classifiers as in **(b)**.

Table S6: **Logistic regression coefficients, mean ± sem.**

(a) **Diagnosis ∼ log perceptual variability**

	coefficient	t value	p value
intercept	13.3 ± 4.8	2.78	0.0055**
log perceptual variability (σ)	4.7 ± 1.7	2.80	0.0051**

(b) **Diagnosis ∼ log TIMO**

	coefficient	t value	p value
intercept	1.9 ± 0.9	2.07	0.038*
log TIMO	1.1 ± 0.5	2.25	0.024*

(c) **Diagnosis ∼ log perceptual variability + log TIMO**

	coefficient	t value	p value
intercept	13.3 ± 4.9	2.70	0.0069**
log perceptual variability (σ)	4.3 ± 1.7	2.47	0.013*
log TIMO	0.63 ± 0.59	1.07	0.28

(d) **Diagnosis ∼ log perceptual variability + log TIMO + log RT median + log RT τ + log lapse rate**

	coefficient	t value	p value
intercept	14.0 ± 5.4	2.57	0.010*
log TIMO	0.49 ± 0.64	0.76	0.44
log RT median	-1.4 ± 1.7	-0.81	0.42
log RT τ	1.4 ± 1.6	0.88	0.38
log perceptual variability (σ)	4.2 ± 1.9	2.23	0.025*
log lapse rate (λ)	0.38 ± 0.37	1.02	0.30