Supplemental Information

Paradigm

The paradigm was a version of the random-dots motion paradigm, where children were told that they needed to help a red rocket that was lost in space by following the stars. Children were asked to point the rocket in the direction of the moving ‘stars’ (dots), by pressing a red (left) or blue (right) button. When a button was pressed (i.e., a choice was made) a small red rocket flashed through the dots in the direction of the button press, accompanied by a flashy sound. Feedback was displayed immediately following the trial: The number of points earned on the preceding trial was displayed in the middle of the screen. To prevent anticipatory responses, the response-stimulus interval (RSI) was varied. On any given trial, the RSI was selected from an exponential distribution with a mean of 0.75 s (range 0.3 - 2.4 s) to ensure stimulus onset was unpredictable and to secure a refractory period. In addition, a penalty delay of 4 s was imposed for responses made within 100 ms of stimulus onset and accompanied by a buzzing error tone (1).

Subjects were first instructed to perform as accurately as possible in order to earn maximum points. Correct choices triggered a rewarding beep. Incorrect choices produced no auditory feedback. After each set of 25 trials, a subtotal score was displayed in a large green font, accompanied by the number of points missed due to incorrect responses in a smaller red font. After each session of 125 trials, subjects had a rest, while the sub-total was displayed on the screen and followed by a picture of the rocket landed on planet, accompanied by an ambient soundtrack. This rest was provided to avoid lapses in attention due to fatigue.

Model fitting procedure

Drift diffusion model (DDM) fitting. The DMAT toolbox maximizes the likelihood of observing a proportion of responses with a given number of reaction time (RT) bins with a multinomial likelihood function (MLF). Here, RT-bins are defined by the 1st, 2nd, 5th, 10th, 30th, 50th, 70th, and 90th quantiles of the RT-distribution. The DDM model fit by the DMAT toolbox consists of seven parameters: three parameters for the decision process (decision threshold a, starting point z and drift rate v), a parameter for non-decision processes (non-decision time T_{er}), and three parameters for intertrial variability (variability in bias s_z, variability in non-decision time s_t, and variability in stimulus quality η, see (2-5). Additionally, two parameters were fit to account for outlier trials (6). We used the DMAT Nelder-Mead SIMPLEX optimization algorithm to fit the DDM. To reduce possible local minima, the algorithm restarts six times performing a number of short runs with 250 iterations, before the final run of 5000 iterations.

Fast guesses. The DMAT toolbox provides a method to account for fast guesses, defined as impulsive choices based on a guess or caused by a distraction. As these data can bias the decision parameters, it is important to correct for such contaminant data (6). Fast guesses are determined by the DMAT toolbox using the exponentially weighted moving average (EWMA) method. This method assumes that fast guesses can be identified by very short RTs with accuracy levels at chance level. In short, all RTs are sorted to find the minimal RT at which accuracy data starts to deviate from chance. All responses faster than this minimal RT are then excluded from further analyses (5, 6).

Most subjects made more fast guesses in the speed than in the accuracy sessions ($F_{1,53} = 66.5$, $P < 0.001$). Subjects with ADHD made more fast guesses than control subjects in both the accuracy and speed sessions (8% vs 16%; $F_{1,53} = 5.74$, $P < 0.05$). Exclusion of fast guesses had a minimal effect on accuracy and RT and did not affect the significance of the between group effects described. When correcting for fast guesses, scores per minute were lower for both groups, especially in the speed sessions. Between group effects on scores per minute did not change after correcting for fast guesses.

Model selection. We fit four different models for each subject separately, to find the best balance between fit-error reduction and power. Each model permitted one or more parameters to vary within individual subject data to detect changes in parameters across speed and accuracy conditions: Model 1 constrained all parameters to be equal across conditions. Model 2 allowed decision threshold and drift rate to vary between, but not within, speed and accuracy conditions. Model 3 allowed decision threshold, drift rate and non-decision time to vary between speed and accuracy conditions and allowed drift-rate to co-vary with motion-strength within these conditions. Model 4 allowed decision threshold, non-decision time and drift rate to vary freely between and within speed and accuracy conditions.

For all models, starting point (z) was fixed at half of the decision threshold (a), as no differences in prior information (bias) were expected. All other parameters were kept constant across conditions, allowing for between, but not within subject comparisons. Previous studies have shown that despite these within-subject restrictions, the DDM model can account for changes in accuracy and RT data due to manipulations in speed and accuracy instructions and difficulty (2).

For each subject and each model, the Akaike information criterion (AIC) and Bayesian information criteria (BIC) were calculated to determine the model with the best trade-off between fit quality and model complexity (5). Results are summarized in Table S2. For both subjects with and without ADHD, mean AIC and BIC scores were lowest for model 3. As such, we chose this model for our further analyses.

Quantile probability plots. For each condition of the experiment the proportion correct choices and RT distributions are plotted in a quantile probability plot. These plots represent all the data, together with the quantile probability functions, showing the diffusion model that describes the data. On the x-axis, conditions are split in the proportion incorrect (left) and correct (right) responses. For each condition the RT distribution is plotted on the y-axis, divided into eight quantiles. As such the plot represents how accuracy changed across different levels of difficulty (coherence) and how speed instructions affected the data (2, 6) (see Figure S2).

The *group* quantile probability plots were generated as follows: First, for each subject, each condition was split into incorrect and correct responses. Then for each of these datasets the proportion correct choices and the 1^{st}, 2^{nd}, 5^{th}, 10^{th}, 30^{th}, 50^{th}, 70^{th}, and 90^{th} quantiles of the RT-distribution were calculated. These were then averaged across subjects (7). To generate the *group* quantile probability functions, we used the DMAT toolbox to generate 1000 simulated data-sets based on the individual DDM parameters for each subject. From each simulated dataset, RT quantiles and proportion correct responses were calculated and averaged within and across subjects.

Additional results

Drift rate & non-decision time. We found higher drift rates for speed compared to accuracy sessions for all subjects, across all difficulty levels (Figure 4B). These results possibly reflect practice effects, showing increased sensitivity to the motion stimulus in speed sessions. Overall drift rates tended to be larger for subjects with ADHD, which was consistent with the higher proportion of moving dots in the motion stimuli for this group (Figure S1). However, between group or interaction effects were not significant ($P > 0.23$).

We found no group effect on non-decision time. However, there was an overall effect of session on non-decision time, with both groups showing shorter non-decision times in the speed sessions. Stressing speed has been shown to impact (pre-)motor responses and post-decision processes, in addition to the decision threshold (8, 9). However, in our findings, impulsive and hyperactive symptoms were not related to differences in non-decision time, but rather to differences in decision threshold.

Variability in decision threshold, drift rate and non-decision time. We used non-parametric Mann-Whitney tests to investigate group differences for parameters that capture inter-trial variability of the main parameters of the drift-diffusion model (s_t, s_z, η; Table S3). The variable s_z usually reflects inter-trial variability at starting point z; the time point where no evidence for either alternative has yet been collected (3, 5). However, in our data starting point z was fixed at $z = a / 2$. As such, any variability in starting point (s_z) is driven by variability in the decision threshold a. There were no differences between groups on these measures, although there was a trend-level effect for s_z ($t_{53} = 1.7$, $P = 0.091$). To test whether permitting within subject-variance in s_z produced a better fit to the data, we performed an exploratory analysis, where s_z was allowed to vary across all conditions between and within speed and accuracy sessions. Mean AIC and BIC values were higher for this model than for the model chosen (model 3). As such, allowing increased variability in this parameter did not lead to a better model.

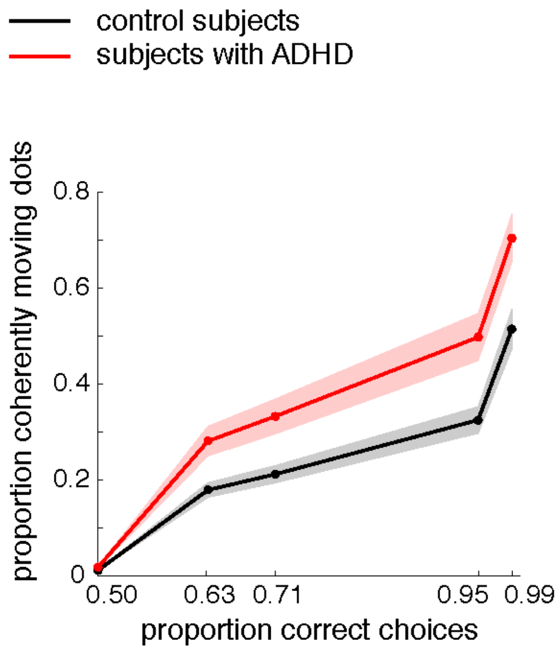

Figure S1. Proportion of coherently moving dots, corresponding to different performance levels in the random dots motion task. For each subject, the proportion of coherently moving dots was obtained corresponding to performance levels of 50%, 63%, 71%, 95% and 99% correct choices. These five difficulty levels were then used in the experimental sessions. Lines represent mean motion coherence. Shaded areas represent standard error of mean (SEM). Overall, subjects with ADHD required a higher proportion of coherently moving dots (red) to obtain the same level of performance as the control subjects (black).

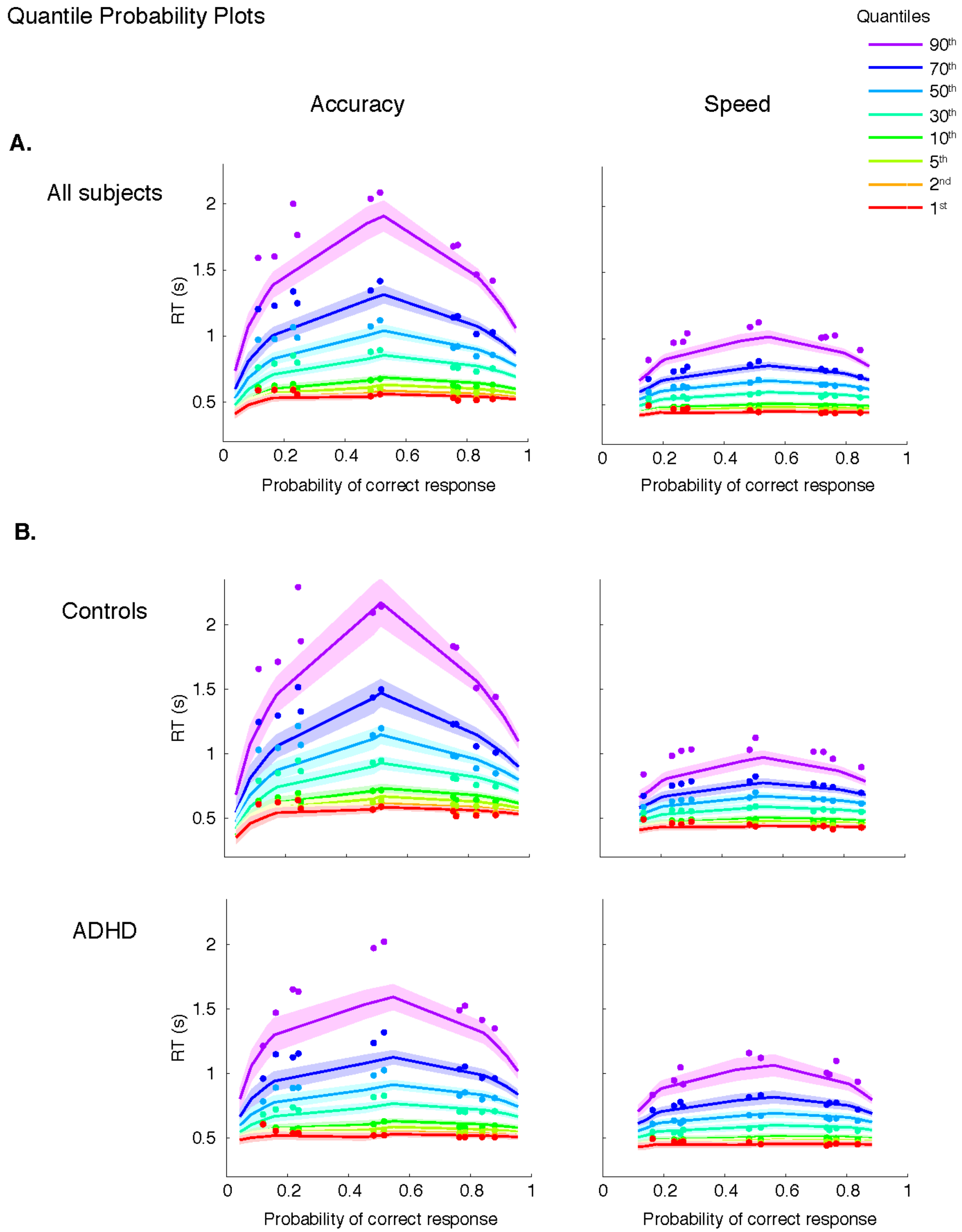

Figure S2. Group quantile probability plots for accuracy (left) and speed (right) sessions. Each graph represents group averages of the proportion correct choices and reaction time (RT) distributions for each difficulty level (datapoints) and the DDM quantile probability functions describing them (lines). RT distributions are represented by eight quantiles (colors), plotted along the y-axis for each condition. Coherence conditions are split into correct and incorrect responses and divided over the x-axis, representing response probability: i.e., if correct responses occurred on 80% of trials within a condition, then correct RT quantiles would be plotted above the point 0.8 on the x-axis, and incorrect RT quantiles from the same condition would appear above the point 0.2 on the x-axis. Lines connecting the quantiles between conditions represent changes in RT distributions across difficulty levels, for incorrect and correct responses. Shaded areas around each line are the standard error of mean (SEM). Distances between quantiles, across the y-axis reflect the effect of decision threshold *a* on RT. Distances and slope of the lines between conditions represent the effect of drift rate *v*. Non-decision time *Ter* determines the placement of RT distributions on the y-axis. **(A)** Group quantile probability plots for all subjects. **(B)** Group quantile probability plots for control subjects and subjects with ADHD. Most graphs show a dramatic change of accuracy across difficulty. The inverted U-shapes indicate that drift rates were changing across conditions (6). Differences between accuracy and speed sessions show a large reduction of RTs, representing the effects of the speed instructions. Although the quantile probability functions sufficiently describe the data, they do deviate from the data at some points. Specifically, fits are worse for the incorrect data and for the higher quantiles in the accuracy sessions. This is possibly due to the lower number of incorrect trials for the accuracy sessions.

Table S1. Descriptive statistics for typically developing controls and subjects with ADHD (mean ± SD).

	Controls (N = 30)	**ADHD (N = 25)**
Gender (M/F)	21/9	22/3
Age	12.9 (4.0)	11.8 (3.1)
Tanner (A) stage	2.7 (1.8)	2.4 (1.7)
IQ	110 (18)	102 (16)
Hand preference (L/R)	3/27	2/23
ADHD - type (number of subjects)		
Inattentive	0	9 **
Hyperactive	0	4 **
Combined	0	12 **
DISC-P – total symptom scores		
Inattentiveness	0.6 (1.3)	6.3 (2.0) **
Hyperactivity/ impulsivity	0.5 (1.2)	5.5 (2.2) **
ODD	0	7 **
TRF (controls/ADHD: N = 20/19)		
ADHD	51.1 (2.2)	60.1 (7.2) **
I	38.8 (19.1)	68.5 (22.0) **
II	38.4 (19.2)	78.3 (23.2) **
ODD	51.2 (2.5)	60.2 (7.1) **
CD	51.2 (3.6)	55.4 (5.1) **
Affective problems	52.5 (4.2)	55.9 (4.8) *
Anxiety problems	52.6 (4.2)	56.6 (6.4) *
Somatic problems	50.4 (1.6)	51.8 (4.3)
Anxious/depressed	53.7 (4.7)	56.6 (6.7)
Withdrawn/depressed	55.7 (7.9)	55.5 (4.9)
Somatic complaints	51.2 (2.8)	51.7 (4.1)
Social problems	52.4 (2.8)	57.8 (6.9) **
Thought problems	51.2 (3.0)	57.8 (6.9) **
Attention problems	50.8 (2.0)	57.9 (6.3) **
Rule-breaking behavior	51.2 (2.9)	54.4 (4.9) *
Aggressive behavior	51.1 (2.8)	60.4 (7.6) **
SES		
Maternal education (years)	13.2 (2.1)	13.7 (1.8)
Paternal education (years)	13.9 (2.3)	13.6 (2.6)
Currently taking stimulant medication	0	6

ADHD, attention deficit hyperactivity disorder; CD, conduct disorder; DISC-P, Diagnostic Interview Schedule for Children, parent version; F, female; L, left; M, male; ODD, oppositional defiant disorder; R, right; SES, socioeconomic status; TRF, Teacher Report Form.

All subjects on stimulant medication discontinued treatment 24 hrs prior to participating in the study.

** $P < 0.01$, * $P < 0.05$.

Table S2. Model selection. Mean values of Akaike information criterion (AIC) and Bayesian information criteria (BIC) across subjects for each group.

	AIC				BIC			
Model	1	2	**3**	4	1	2	**3**	4
Controls (N = 30)	2267.7 (366.2)	2168.7 (333.5)	**2066.5 (345.1)**	2078.5 (356.2)	2300.3 (366.8)	2209.4 (334.4)	**2115.2 (346.1)**	2216.6 (359.8)
ADHD (N = 25)	1906.7 (463.4)	1852.8 (441.5)	**1769.2 (430.4)**	1791.0 (428.7)	1938.4 (464.9)	1892.2 (443.3)	**1816.4 (432.6)**	1923.9 (436.1)

AIC and BIC values were calculated for each model to determine which model had the best tradeoff between fit quality and model complexity. Model 3 had the lowest AIC and BIC values, indicating the best fit to the data.

Table S3. Mean (SD) values of drift diffusion model (DDM) parameters for typically developing controls and subjects with ADHD reflecting variability of main parameters decision threshold, drift rate and non-decision time.

DDM parameter		**controls (N = 30)**	**ADHD (N = 25)**
variability in a	s_z	0.023 (0.04)	0.048 (0.06)
variability in T_{er}	s_t	0.240 (0.17)	0.205 (0.16)
variability in v	η	0.048 (0.09)	0.064 (0.08)

1. Simen P, Contreras D, Buck C, Hu P, Holmes P, Cohen JD (2009): Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *J Exp Psychol Hum Percept Perform*. 35:1885-1897.
2. Ratcliff R, McKoon G (2008): The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput*. 20:873-922.
3. Ratcliff R (1978): A theory of memory retrieval. *Psychol Rev*. 85:59-108.
4. Vandekerckhove J, Tuerlinckx F (2008): Diffusion model analysis with MATLAB: A DMAT primer. *Behav Res Methods*. 40:61-72.
5. Vandekerckhove J, Tuerlinckx F (2007): Fitting the Ratcliff diffusion model to experimental data. *Psychon Bull Rev*. 14:1011-1026.
6. Ratcliff R, Tuerlinckx F (2002): Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychon Bull Rev*. 9:438-481.
7. Ratcliff R (1979): Group reaction time distributions and an analysis of distribution statistics. *Psychol Bull*. 86:446-461.
8. Rinkenauer G, Osman A, Ulrich R, Muller-Gethmann H, Mattes S (2004): On the locus of speed-accuracy trade-off in reaction time: inferences from the lateralized readiness potential. *J Exp Psychol Gen*. 133:261-282.
9. Voss A, Rothermund K, Voss J (2004): Interpreting the parameters of the diffusion model: an empirical validation. *Mem Cognit*. 32:1206-1220.