

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <http://orca.cf.ac.uk/115941/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Powell, Georgie, Jones, Catherine R., Hedge, Craig, Charman, Tony, Happe, Francesca, Simonoff, Emily and Sumner, Petroc 2019. Face processing in autism spectrum disorder re-evaluated through diffusion models. *Neuropsychology* 33 (4) , pp. 445-461. 10.1037/neu0000524 file

Publishers page: <http://dx.doi.org/10.1037/neu0000524> <<http://dx.doi.org/10.1037/neu0000524>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



## Title

Face processing in autism spectrum disorder re-evaluated through diffusion models

## Authors

Georgie Powell<sup>1\*</sup>, Catherine R Jones<sup>1</sup>, Craig Hedge<sup>1</sup>, Tony Charman<sup>2</sup> Francesca Happé<sup>3</sup>, Emily Simonoff<sup>4</sup>, Petroc Sumner<sup>1</sup>.

<sup>1</sup> School of Psychology, Cardiff University, Park Place, Cardiff, CF10 3AT, UK.

<sup>2</sup> Department of Psychology, King's College London, Institute of Psychiatry Psychology and Neuroscience, London, UK

<sup>3</sup> SGDP Research Centre King's College London, Institute of Psychiatry Psychology and Neuroscience, London, UK

<sup>4</sup> Department of Child & Adolescent Psychiatry and NIHR Biomedical Research Centre for Mental Health, King's College London, Institute of Psychiatry Psychology and Neuroscience, London, UK

\* *Corresponding author*

## Acknowledgements

We are grateful to the adolescents and families who took part in this study. We thank Bob Joseph for sharing his stimuli used in the face recognition task (see Joseph & Tanaka, 2003) and Kate Lawrence (Elgar) and David Skuse for sharing their stimuli included in the egocentric eyegaze task (see Elgar et al., 2002). We additionally thank Jeni Tregay, Anita Marsden and Rebecca Phillips for help with data collection. Both experimental tasks were realised using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience. We also thank the ESRC for funding this project (ES/K002325/1) and to the Medical Research Council (G0400065) for funding the original study.

## Abstract

**Objective:** Research using cognitive or perceptual tasks in autism spectrum disorder (ASD) often relies on mean RT and accuracy derived from alternative-forced choice paradigms. However, these measures can confound differences in task-related processing efficiency with caution (i.e. preference for speed or accuracy). We examined whether computational models of decision-making allow these components to be isolated.

**Method:** Using data from two face-processing tasks (face recognition and egocentric eye-gaze discrimination), we explored whether adolescents with ASD and wide ranging intellectual ability differed from an age and IQ matched comparison group on model parameters that are thought to represent processing efficiency, caution, and perceptual encoding/motor output speed.

**Results:** We found evidence that autistic adolescents had lower processing efficiency and caution, but did not differ from non-autistic adolescents in the time devoted to perceptual encoding/motor output. These results were more consistent across tasks when we only analysed participants with IQ above 85. Cross-task correlations suggested that processing efficiency and caution parameters were relatively stable across individuals and tasks. Furthermore, logistic classification with model parameters improved discrimination between individuals with and without ASD relative to classification using mean RT and accuracy. Finally, previous research has found that ADHD symptoms are associated with lower processing efficiency, and we observed a similar relationship in our sample, but only for autistic adolescents.

**Conclusions:** Together, these results suggest that models of decision-making could provide both better discriminability between autistic and non-autistic individuals on cognitive tasks and also a more specific understanding of the underlying mechanisms driving these differences.

## Keywords

Autism spectrum disorders, ADHD, decision-making, drift-diffusion model, face processing

## Public significance statement

Researchers are often interested in how individuals with autism spectrum disorders (ASD) differ from those without on cognitive and perceptual tasks. We examined a different way to analyse results from these tasks, using computational models, and found that they better described differences between adolescents with and without ASD than standard measures. One potential application of these findings is the development of more targeted training paradigms in the future.

## Introduction

Autism spectrum disorder (ASD) is characterised by social communication difficulties and the presence of restrictive and repetitive behaviours (American Psychiatric Association, 2013). In addition to these core symptoms, there has been enduring interest in the cognitive and perceptual phenotype in ASD and associated neurocognitive theories of ASD (e.g. Happe & Ronald, 2008; Pellicano & Burr, 2012 ). These theories tend to suggest that different cognitive processing styles can impact the development of social and communication deficits characteristic of ASD. The search for cognitive and perceptual differences in ASD has been comprehensive and spans a wide array of tasks. However, in many areas there is still no clear consensus on which tasks best differentiate autistic individuals from those without, and different studies often report conflicting findings (e.g. Charman et al., 2011; Kuiper, Verhoeven, & Geurts, 2016; Leekam, 2016; Simmons et al., 2009).

Most cognitive and perceptual tasks ask participants to make decisions – often in the form of alternative-forced-choice paradigms. For example, ‘Are the dots moving to the left or the right?’ or ‘Is this face old or new?’. The outcome measures are usually whether the participant made the correct choice (accuracy) and how long it took them to make it (response time, RT). However, analysing RT and accuracy separately can confound differences in caution (i.e. speed/accuracy preference) with task-related ‘processing efficiency’ (i.e. directly related to task proficiency). An individual who is very cautious, or has a preference for accuracy over speed, will tend to make few errors but will have long RTs. While an individual who is very proficient at a task, and processes information efficiently, will tend to be both faster and more accurate. Thus, differences in RT and accuracy between groups with and without ASD could potentially arise due to differences in caution, task-related processing efficiency, or some mixture of the two. Computational models of decision making allow these two components to be disentangled (for reviews see Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff, Smith, Brown, & McKoon, 2016; Teodorescu & Usher, 2013), and may therefore be both more sensitive to task differences between individuals with and without ASD, and more diagnostic as to what these differences represent in terms of underlying mechanisms. Using data from two face processing tasks, we investigated whether parameters derived from computational models of decision-making are better able to differentiate between individuals with and without ASD than traditional behavioural measures of accuracy and RT. We were also interested in whether any group differences in model parameters representing caution and processing efficiency are consistent across tasks or whether they are task-specific.

*Traditional behavioural measures versus models of decision making*

Reporting RT and accuracy separately can result in scenarios where, hypothetically, a group of autistic individuals could respond more quickly on a task than those without ASD but are not as accurate. What could a researcher conclude about the relative performance on this task? Such a result could occur because accuracy and response time confound efficiency of information processing (i.e. directly related to task proficiency) with caution (i.e. preference for speed or accuracy). Accumulator models such as the drift diffusion model (DDM, Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff et al., 2016) combine data from accuracy and RT, and allow these components to be isolated. In the example above, the DDM might reveal that autistic individuals are less cautious than individuals without, resulting in faster but less accurate responses, but task-related efficiency of information processing is similar between the groups.

The DDM assumes that when presented with a choice between two alternatives, information accumulates stochastically until one of two response boundaries is reached (Figure 1). The time it takes to reach a boundary corresponds to the RT and which boundary is reached corresponds to the choice (e.g. correct/incorrect). The rate of accumulation, or 'drift-rate', is thought to represent the quality of information extracted from the stimulus over time or the 'processing efficiency'. Therefore, higher drift-rates result in faster and more accurate responses – overall more 'proficient' performance on the task. Difficult tasks or conditions result in lower drift-rates. Within tasks, individuals can differ in their drift rates, and indeed, this may be closely related to stable traits such as IQ (Ratcliff, Thapar, & McKoon, 2011; van Ravenzwaaij, Brown, & Wagenmakers, 2011). Within the DDM, caution, or preference for speed/accuracy, is represented by the degree of boundary separation. Very cautious individuals will have widely separated boundaries and so require more information to accumulate before making a decision. This means that the decisions they make will tend to be more accurate but relatively slow. Whereas, less cautious individuals have boundaries closer together, resulting in faster responses that tend to be less accurate. The DDM also assumes that a component of the overall response time is devoted to perceptual encoding of the stimulus and motor output. This is represented by a parameter called 'non-decision time', which is added to the result of the accumulation process to yield the overall RT.

Besides a means to interpret RT and accuracy together, models like the DDM aim to fit the shape of the whole RT distribution and so may capture additional elements of the data that are missed by relying on only average RT. Recently, this has been successfully demonstrated in the literature on attention deficit hyperactivity disorder (ADHD), a condition characterised by impulsivity, inattention and hyperactivity (American Psychiatric Association, 2013). Several papers have reported that individuals with ADHD show greater intra-participant variability in RT distributions (Castellanos et al., 2005; Johnson et al., 2007; Klein, Wendling, Huettner, Ruder, &

Peper, 2006; Kofler et al., 2013; Kuntsi, Oosterlaan, & Stevenson, 2001; Mullins, Bellgrove, Gill, & Robertson, 2005; Russell et al., 2006). In particular, individuals with ADHD have RT distributions with elongated tails, which when fitted with an ex-Gaussian function, is revealed in larger values of the exponential component 'tau' (Epstein et al., 2011; Geurts et al., 2008; Karalunas, Huang-Pollock, & Nigg, 2013; Kofler et al., 2013; Leth-Steensen, Elbaz, & Douglas, 2000). Analysis of data with the DDM has revealed lower drift-rates in individuals with ADHD, which could explain this increase in intra-participant variability (Karalunas, Huang-Pollock, & Nigg, 2012; Karalunas et al., 2013; Metin et al., 2013; Salum et al., 2014). This suggests that individuals with ADHD may not process information as efficiently as neuro-typical controls. Although one might predict that individuals with ADHD would be more impulsive (i.e. less cautious) than individuals without ADHD, analysis with the DDM found caution to be similar between groups. Distinguishing these processes is crucial for developing effective cognitive training therapies that target key areas of difficulty (Abikoff, 1991). The DDM has been used to successfully model, and reveal differences in, decision making processes in a number of other subpopulations, such as older adults (Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2004) and individuals with Parkinson's disease (Zhang et al., 2016) and schizophrenia (Moustafa et al., 2015).

#### *Drift-diffusion model and ASD*

There is evidence that autistic individuals, like those with ADHD, show increased intra-participant variability and more elongated RT distributions than neuro-typical controls on a range of cognitive tasks (e.g. Geurts et al., 2008; for a meta-analysis see Karalunas, Geurts, Konrad, Bender, & Nigg, 2014). This could indicate a global impairment in processing efficiency that is not modality or task specific. This might in turn explain the prevalence of conflicting findings in the cognitive and perceptual literature on ASD, because intra-participant variability could reduce test-retest reliability (Milne, 2011).

On the assumption that intra-participant variability is explained in ADHD by lower drift-rates, we could predict that the elevated intra-participant variability in ASD would also be associated with lower drift-rates. Only two studies have used the DDM to explore differences in ASD. The first found that autistic adults did not differ in drift-rate from non-autistic adults when tested on an orientation discrimination task (Pirrone, Dickinson, Gomez, Stafford, & Milne, 2016). Rather, this study suggested that autistic individuals have wider boundary separation (i.e. are more cautious) and longer non-decision times than non-autistic individuals. The second study examined autistic children and also found wider boundary separations, this time using the go trials from the Stop Signal task (Karalunas et al., 2018). However, this study did find slower drift-rates in autistic participants. Given

the large overlap between ASD and ADHD in both genotype and phenotype (e.g. Gargaro, Rinehart, Bradshaw, Tonge, & Sheppard, 2011; Keen & Ward, 2004; Leyfer et al., 2006; Mayes, Calhoun, Mayes, & Molitoris, 2012; Rommelse, Franke, Geurts, Hartman, & Buitelaar, 2010; Simonoff et al., 2008), slower drift-rates in the ASD group could be attributable to the inclusion of individuals with increased ADHD traits or comorbidity (Karalunas et al., 2014). However, Karalunas et al. (2018) found that differences in drift-rate remained when ADHD symptoms were controlled for.

### *Present study*

In the present study, our aim was to investigate whether autistic adolescents differed from those without ASD on DDM parameters of drift-rate (processing efficiency), boundary separation (caution) and non-decision time (perceptual encoding/motor output). We used the DDM to analyse data from wave two of the Special Needs and Autism Project (SNAP, Baird et al., 2006), which explored the cognitive phenotype of autistic adolescents (see Charman et al., 2011). The participant sample in SNAP comprises individuals with ASD and a matched comparison group, and deliberately covers a broad range of intellectual ability (IQ range: 52-133). Our main analysis contained the whole sample of participants, but we also repeated all analyses only on individuals with full-scale IQ scores above 85. This was to ensure that any results we reported were not being driven by individuals with low IQ, additional neurodevelopmental conditions and special educational needs.

Two tasks, face recognition and egocentric eye-gaze discrimination, were suitable for modelling with the DDM because they required participants to make alternative forced-choice decisions that were not presented using a staircase procedure. In order to maximise trial numbers for model fitting, and because our main area of interest was in global differences in behavioural measures and parameters, we collapsed across conditions within task. Autistic individuals are known to show deficits in face recognition, which has been attributed to a tendency to focus on the mouth relative to the eyes and a reduction in holistic processing (Joseph & Tanaka, 2003; Langdell, 1978). There is also evidence that autistic individuals show deficits in eye-gaze discrimination, although findings are somewhat mixed and may be depend on factors such as the direction of the head, the age and gender of participants, and whether the task involves spontaneous monitoring of gaze (Ashwin, Ricciardelli, & Baron-Cohen, 2009; Forgeot d'Arc et al., 2017; Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Webster & Potter, 2008). Autistic individuals might show deficits in face processing tasks because they find such stimuli more aversive than non-autistic individuals (e.g. Tottenham et al., 2013).

Taken together, autistic participants would be expected to perform less well on both face recognition and egocentric eye-gaze discrimination than the comparison participants without ASD. For the traditional behaviour measures, this could affect both accuracy and RT, although as discussed above, the exact pattern of results is conflated with participants' unique preference for speed or accuracy. In terms of model parameters, a deficit in task performance would be reflected in reduced drift-rate (efficiency of information processing). Since drift rate is explicitly separated from caution (boundary separation) by the model, it ought to be a purer measure of task processing deficit than RT or accuracy rates. Therefore, a key question for this study is whether drift rate differs between groups in either or both tasks. The corresponding question is whether boundary separation instead accounts for any behavioural differences between groups. Non-decision time is not expected to differ across groups, but could potentially do so due to low level perceptual differences.

Drift-rate might also be reduced in the ASD group because of increased intra-participant variability, which is often captured in the DDM as a reduction in drift-rate. As discussed above, the relationship between intra-trial variability, drift-rate and ASD might be moderated by comorbid ADHD, and so we also explored the relationship between model parameters and ADHD symptoms using a parent-report questionnaire.

If model parameters do represent stable character traits, we would predict that group differences in these parameters are stable across tasks and the parameters themselves would correlate across tasks.

Lastly, we predicted that DDM parameters would be better able to distinguish autistic and non-autistic individuals than standard behavioural measures (mean RT/accuracy), because they are able to separate variance due to caution (i.e. speed-accuracy trade-off) from task or individual specific effects. In order to investigate this, we trained a logistic classifier on each measure and compared discriminability performance (inspired by Zhang et al., 2016). The purpose of this was not to identify a measure that is diagnostically relevant for ASD, but rather to test whether the DDM parameters offer better discriminability than traditional measures when searching for task or trait specific differences between groups.

## Methods

### *Participants*

The full sample of participants consisted of 75 autistic adolescents (mean age = 15.5 years, SD = 0.54, 67 male) and 46 non-autistic adolescents (mean age = 15.5 years, SD = 1.3, 45 male). The participants were drawn from a cohort of 100 autistic adolescents and 57 non-autistic adolescents who were part of wave two of the Special Needs and Autism project (Charman et al., 2011). Adolescents were excluded if data were not available on both tasks. We also excluded participants who scored below chance on either tasks or whose average RT exceeded 3 x median absolute deviation on either task. Thirty-three of the participants excluded had ASD, 25 were male, the average full scale IQ was 71.7 (SD = 19.5), the average age was 15.4 years (SD = 0.4) and the average ADHD symptom score was 6.3 (SD = 2.3). Consensus clinical ICD-10 diagnosis of ASD was confirmed using the Autism Diagnostic Interview - Revised (Lord, Rutter, & Le Couteur, 1994) and Autism Diagnostic Observation Schedule – Generic (Lord et al., 2000), alongside measures of IQ, language and adaptive behaviour (see Baird et al., 2006).

Participants in the comparison group comprised 18 children from wave one without ASD but with a range of primary ICD-10 diagnoses (8 with mild mental retardation, 3 moderate mental retardation, 2 ADHD, 2 specific reading and spelling disorder, 1 expressive/receptive language disorder, 1 no primary diagnosis), and 28 typically developing adolescents recruited from mainstream schools specifically for wave two. By parent report, none of the typically developing adolescents had a psychiatric or developmental diagnosis, a statement of special educational needs, or were receiving medication. Parents completed the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) for 23 of the typically developing participants and none scored 15 or above, the cut-off for ASD (mean score = 3.39, SD = 3.63). IQ was measured using the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). Bayesian independent t-tests using default priors suggested moderate evidence that there was no difference between the groups on age (BF01 = 5.0) or IQ (BF01 = 3.9).

All analyses were repeated for the subgroup of participants with full scale IQ scores above 85, which consisted of 45 autistic adolescents and 30 non-autistic adolescents. We selected a cut-off of 85 because it is one standard deviation below mean IQ, and it excluded all of the comparison group participants who had ICD-10 diagnoses relevant to the task, while critically allowing the groups to remain largely IQ matched. Two individuals in the comparison group had an ICD-10 diagnosis of Specific Reading Impairment, but the tasks did not involve reading. Results with these two participants removed are shown in Supplementary Materials (Figure 3-5). The same exclusion

criteria detailed above were also applied to this subgroup. Bayesian independent t-tests suggested moderate evidence in favour of no age difference between the groups ( $BF_{01} = 4.1$ ), and no evidence either way for a difference in full Scale IQ between the groups ( $BF_{01} = 0.47$ ). See Table 1 for participant demographics and psychological measurements. Approval for the study was granted by the South East Research Ethics Committee (05/MRE01/67) and informed consent was obtained from all participants

### *Tasks*

Most participants were tested in a quiet testing space within the researcher's institution, but a minority were tested at school or home in a 1:1 quiet environment. The tasks were part of a larger battery of 58 tasks that were administered over two separate testing sessions, lasting 3 to 3.5 hours each excluding breaks. The tasks most appropriate for modelling with the DDM were a face identity recognition task and an egocentric eye-gaze discrimination task. Both tasks were programmed in Matlab v6.5 (Mathworks Inc., Sherbon, MA) using Cogent 2000 (Wellcome Department of Imaging Neuroscience, UCL Institute of Neurology, London, UK; <http://www.vislab.ucl.ac.uk/cogent.php>) and presented on a Hewlett-Packard laptop with a 15" LCD display screen. In order to maximise the number of trials for fitting, we collapsed RT and accuracy data across conditions. This was justified because our main area of interest was whether the DDM parameters added anything to general differences in RT and accuracy that might be consistent across tasks, rather than condition-specific differences between groups.

### *Face recognition*

This task was based on Joseph and Tanaka's (2003) assessment of holistic and part based face recognition in autistic children. In the task, recognition of faces and parts of faces (mouth or eyes) was measured. Additionally, whether participants were prompted or unprompted to look at the mouth or eyes during the study (encoding) phase was manipulated.

Testing began with a practice phase where participants were told they would be playing a game in which they would see a picture in the middle of the screen. After presentation of the picture they were then presented with the initial picture alongside a foil. They were told to, 'Choose the one that is the same'. The first trial used photographs of cars and the second used photographs of faces. Images were presented in PowerPoint. Participants needed to pass both trials to progress to the testing phase.

Faces used in the testing phase were created by Joseph and Tanaka (2003). They were greyscale digital images of 12 children's faces (6 boys, 6 girls) with a neutral expression. Their size was approximately 10 x 14 cm and they were set against a grey background. Each image was a digitally manipulated composite created from original digital photographs of different children (see Joseph and Tanaka for details). Within each condition (prompted or unprompted) each target face was presented twice. One was a 'whole' presentation in which the whole target face was followed by paired presentation of the whole target face and a whole foil face. The other was a 'part' presentation, in which the whole target face was followed by paired presentation of part of the target face and part of a foil face. The part was either the eyes or mouth. Foils were the target faces with either the eyes or mouth replaced with those from unused photographs. In total there were 48 pairs of images (12 targets x 2 foil types (mouth or eyes altered) x 2 image types (whole face or part face), which were divided into two sets of 24 (prompted and unprompted conditions). Stimuli presentation was in one of six pseudo-random orders.

In the first part of the trial, the target image was presented in the centre of the screen for 3.5s (study phase). Immediately following this, the target and foil pair were presented. Participants were asked to decide which image was the same as the one they had just seen, using a keypress. Whether the target face was presented on the left or right was counterbalanced across trials. After the participant had made a response, or following 7 s if they failed to respond, the next trial was initiated. For the unprompted trials, the inter-trial interval was 1s, and consisted of the presentation of a white rectangle (500 ms) followed by a blank screen (500 ms). For the prompted trials, the inter-trial interval was also 1s but contained the instruction to 'look at the eyes' or 'look at the mouth', presented in white font (Arial, 70) on a black background. Before the prompted condition the participants were told, "You will either be told to look at the eyes or look at the mouth; this is a clue to help you". The order of the trials was pseudo-randomised, with 6 different orders used. Viewing distance for the task was approximately 75 cm, in line with Joseph and Tanaka (2003). Accuracy and RT were recorded for each trial.

#### *Egocentric eye-gaze discrimination*

This task was based on Elgar, Campbell and Skuse (2002) and measured discrimination of eye-gaze direction (straight ahead vs averted) for different head angles (straight ahead vs averted).

Participants first completed three practice trials to ensure that the task was understood. The first trial had a face that looked straight ahead and the second and third had faces that were angled to the left or right. Participants had to complete all three trials successfully to take part in the experiment. Practice trials were presented in PowerPoint.

Stimuli used in this experiment were provided by Elgar et al. (2002) and were colour photographs of a single adult female, sized approximately 10 cm x 15 cm. The female was captured with her head facing straight ahead (12 trials), or turned at a 10° angle to the left (12 trials) or right (12 trials). For each of these head positions, for 4 trials the eye gaze was straight ahead, for 4 trials the eye gaze was to the right, and for 4 trials the eye gaze was to the left. The position of the left and right eye gaze was at a 10° angle from the centre (head angled conditions) or at a 5° or 10° angle from the centre (head straight ahead condition). Trials were presented in one of six pseudo random orders.

For each trial, the stimulus was presented in the centre of the screen for up to 7 s. If the participant did not respond within 7 s then the trial was terminated. A black background was used throughout the experiment. Each trial was preceded by a white fixation asterisk (1s) followed by a blank screen (500 ms). The participants were told that they had to decide if the person was looking straight at them, or to their left or right and indicate this with a key press. Participants were also instructed to respond as quickly and as accurately as they could. Participants sat approximately 50 cm from the screen, following Elgar et al., (2002). Accuracy and RT were recorded for each trial.

#### *Parent-reported ADHD symptoms*

ADHD symptoms were measured using the parent version of the Strengths and Difficulties Questionnaire (SDQ, Goodman, 1997). The SDQ is a commonly used screening measure for child psychiatric problems. It contains 25 items each scored 0 to 2 that map onto five subscales (hyperactivity-inattention, emotional problems, conduct problems, peer relationship problems and prosocial behaviour) and a total difficulties score. For the present study we only used scores from the hyperactivity-inattention subscale. This subscale includes two items about inattention, two about hyperactivity and one about impulsiveness, which together reflect the core symptoms of ADHD (American Psychiatric Association, 2013).

#### *Hierarchical drift-diffusion model*

Parameters were estimated using the hierarchical drift diffusion model toolbox (HDDM, Wiecki, Sofer, & Frank, 2013) and RT and accuracy data from both tasks. The HDDM uses Bayesian inference procedures to simultaneously estimate parameters at a group level (ASD vs non ASD) and at an individual level. Bayesian fitting methods are particularly advantageous in situations where there are relatively few trials per participant (Ratcliff & Childers, 2015), as was the case in the present

study. This is because trials across participants contribute to the estimate of group-level parameters, which in turn constrain the distribution (likely values) of individuals' parameters.

The HDDM uses the standard analytic solution to the DDM provided by Wald (1947) and Feller (1968):

$$f(x|v, a, z) = \pi a^2 \exp(-vaz - v^2 x^2) \times \sum_k = 1 \infty k \exp(-k^2 \pi^2 x^2 a^2) \sin(k\pi z)$$

Where  $v$  = drift-rate,  $a$  = boundary,  $z$  = bias. It also models the extension to the DDM that includes inter-trial variability in drift-rate, non-decision time and starting point, which was suggested by Ratcliff and Rouder (1998). Extended details about the fitting process and assumed parameter distributions can be found in Wiecki et al. (2013), in the section titled 'Hierarchical Bayesian estimation of the drift-diffusion model'.

A graphical representation of the model is shown in figure 2. We allowed drift-rate boundary separation and non-decision time to vary across individual and group. Starting point bias, and variability in non-decision time, drift-rate and bias were fitted to the overall dataset and were therefore consistent across group and individual. Attempts to vary these resulted in poor convergence due to the relatively small number of trials.

We generated 20000 samples from the posterior distribution for all model parameters using a Markov Chain Monte-Carlo method. To prevent the initial starting values from biasing the final convergence, we discarded the first 2000 samples as burn-in. We only kept every fifth sample in order to minimise auto-correlations between consecutive draws from the posterior distribution ('thinning'). We also assumed a trial outlier rate of 5%. See Wiecki et al., (2013) for more information. Chain convergence was checked using the Gelman-Rubin R-hat statistic (Gelman & Rubin, 1992), which was less than 1.05 for all modals suggesting good convergence.

### *Group classification*

We trained a logistic classifier to distinguish between the ASD and non-ASD groups using either model parameters (drift-rate and boundary) or traditional measures (RT and accuracy). We excluded non-decision time because we did not find evidence of a group difference and it allowed the classification space to contain the same number of features for both behavioural measures and model parameters. A 10 fold cross validation procedure was used. Data were split into 10 subsamples, 9 of which were initially used for training and one for validation. This routine was then

repeated until all the 10 sub-samples had been used for validation. Results were then averaged across the 10 folds. Our performance measure for classification was the receiver operating characteristics (ROC) area under the curve (AUC). The classifier was implemented in WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>). It is worth noting that leave-one-out cross validation procedures can yield higher predication accuracy values than methods using independent test and validation samples, however, our principal concern here was the comparison in discrimination ability between behavioural measures and model parameters.

## Results

### *Results analysis*

Behavioural results were analysed using Bayesian Statistics (JASP-Team, 2016; Rouder, Morey, Speckman, & Province, 2012) and we report associated Bayes Factors which represent the degree to which the data favour the alternative hypothesis (H1) relative to the null hypotheses (H0). Where BF10 indicates evidence towards H1 against H0 and BF01 indicates evidence towards H0 against H1. We use common evaluation criteria, where Bayes Factors below 3 indicate inconclusive evidence, above 3 indicate moderate evidence, above 10 indicate strong evidence, and above 30 indicate very strong evidence (Jeffreys, 1961). Default JASP priors were used throughout and are shown in Table 2. Modal parameters were compared across groups using posterior probability comparisons, which calculate the proportion of posteriors of each parameter that overlap across the two groups. Values here are reported as the proportion of posteriors of the comparison group that are smaller than that of the ASD group. Smaller values therefore indicate more of a difference between the groups.

### *Behavioural results*

Figure 3 shows average accuracy (A) and reaction time (B) for ASD and non-ASD groups for both the face recognition and egocentric eye-gaze discrimination tasks. Two Bayes Factor ANOVAs were performed separately on accuracy and RT data. Both analyses revealed inconclusive evidence of differences between the groups on either task (main effect of group BFs all  $<3$  and  $>0.3$ , supporting neither null or alternative hypothesis) and evidence of no interaction between task and group (BF01 all  $>3$ ). When repeating these analyses for participants with IQs above 85, we found evidence that the non-ASD groups was more accurate than the ASD group (BF10 = 84), but evidence in favour of no

difference between the group in RT ( $BF_{01} = 4.5$ ). Average accuracy and reaction times are shown in Figure 4 (A-B). Descriptive results are shown in Supplementary Tables 1 and 2.

### *DDM parameters*

Posterior parameter estimates of drift-rate (processing efficiency), boundary separation (caution) and non-decision time (perceptual encoding/motor output) are shown in Figure 3C-E (whole sample) and Figure 4C-E (IQ above 85). Posterior probability comparisons indicated that, for the face recognition task, the ASD group showed smaller boundary separation than the comparison group ( $P(\text{comparison } a < \text{ASD } a) = 0.003$ ). There was less evidence supporting a difference in drift-rate or non-decision time ( $P(\text{comparison } v < \text{ASD } v) = 0.3$ ,  $P(\text{comparison } t < \text{ASD } t) = 0.15$ ). For participants with IQ above 85, the ASD group again showed smaller boundary separation than the comparison group ( $P(\text{comparison } a < \text{ASD } a) = 0.001$ ). However, we now additionally found evidence of lower drift-rates in the ASD group ( $P(\text{comparison } v < \text{ASD } v) = 0.021$ ). Non-decision time remained comparable between groups ( $P(\text{comparison } t < \text{ASD } t) = 0.53$ ).

For the eye-gaze task, we found some evidence that the ASD group have lower drift-rate than the comparison group ( $P(\text{comparison } v < \text{ASD } v) = 0.05$ ) but no obvious difference in boundary or non-decision time ( $P(\text{comparison } a < \text{ASD } a) = 0.18$ ,  $P(\text{comparison } t < \text{ASD } t) = 0.83$ ). For participants with IQ above 85, the difference in drift-rate between the groups remained ( $P(\text{comparison } v < \text{ASD } v) = 0.002$ ) and there was still little difference in non-decision time ( $P(\text{comparison } t < \text{ASD } t) = 0.1$ ). However, we now found some evidence that, like the face processing task, participants with ASD had smaller boundary separation ( $P(\text{comparison } a < \text{ASD } a) = 0.05$ ).

Taken together, these results suggest that for participants with IQ above 85, adolescents in the ASD group process information less efficiently and are less cautious than non-autistic adolescents, but show similar non-decision times. When the whole sample of participants across IQ is analysed, differences in information processing and caution are less consistent and potentially more task specific.

### *Model fits*

To ensure that the estimated model parameters were a good representation of the data, we performed posterior predictive checks by simulating 500 experimental data sets for each participant on each task based on their individual parameter estimates. Figure 5 and 6 shows comparisons between observed data and simulated data RT distributions and Table 3 and 4 shows mean square

errors (MSEs). Overall, there is a fairly good correspondence for RTs for both correct and incorrect trials, suggesting that the estimated parameters captured the data well.

### *Group classification*

Our next question was whether DDM parameters are better able to differentiate between individuals with and without ASD than traditional measures of RT and accuracy. To investigate, we trained a logistic classifier to distinguish between the groups using a 10-fold cross validation procedure (see Methods). We first used the mean RT and accuracy data from both tasks as the feature space for classification. This model was able to classify the groups with an AUC value of 0.63. The associated ROC curve is shown in Figure 3F, grey line. We then used the estimated model parameters (drift-rate and boundary separation) across both tasks as the classification feature space. This produced an AUC value of 0.72 (Fig. 3F, black line). Following the method described in Hanley and McNeil (1983), the difference between the AUC values was statistically significant ( $Z = 2.96, p < 0.01$ ). The classification procedure was repeated on participants with IQs above 85 and discrimination ability for the model parameters increased to an AUC of 0.89. Classification using traditional measures also increased in this subset of participants, but to a lesser extent (AUC = 0.71). ROC curves are shown in Figure 4F. The difference between these AUC values was also statistically significant ( $Z = 2.36, p < 0.05$ ). Therefore, the DDM parameters were better able to differentiate between the groups than the traditional behavioural measures for these tasks.

### *Relationship between drift-rate and hyperactivity-inattention*

Previous research has suggested that individuals with ADHD process information less efficiently, i.e. have lower drift-rates, than individuals without ADHD. We were therefore interested in whether individual differences in drift-rates were related to ADHD symptoms in our sample, which were measured using the hyperactivity-inattention subscale from the SDQ. Figure 7 shows scatterplots of these relationships, split by group and task. A moderation analysis was carried out separately for each task, with drift-rate as the IV, group as the moderator and ADHD symptoms as the DV. For the face recognition task, we found evidence that drift-rate negatively predicted ADHD symptoms (BF10 for inclusion of main effects model = 9.4). However, we also found evidence that this relationship was moderated by group (BF10 for inclusion of interaction model = 9.3), whereby drift-rate only predicted ADHD symptoms in the non-ASD group, not the ASD group. This pattern of results was replicated for the egocentric eye-gaze task (BF10 main effect of drift-rate = 4.6, BF10 for interaction = 17.8). However, when IQ was also included in the regression, the data was inconclusive for a main effect of drift-rate or a group x drift-rate interaction on either task (face recognition, main effect of

drift-rate  $BF_{10} = 0.60$ , interaction  $BF_{10} = 1.3$ ; egocentric eye-gaze, main effect of drift-rate  $BF_{10} = 0.45$ , interaction  $BF_{10} = 1.5$ ). This pattern of results was replicated for both verbal and performance IQ, when these were analysed separately.

These results suggest that ADHD symptoms were not related to drift-rate (processing efficiency) in autistic participants. In contrast, in non-autistic participants, drift-rate was negatively associated with ADHD symptoms, but this relationship was strongly influenced by differences in IQ. This is consistent with previous research where drift-rate is often closely related to IQ scores (Ratcliff et al., 2011; van Ravenzwaaij et al., 2011) and indeed a moderation analysis revealed this relationship was present in our study, although only for non-autistic adolescents ( face recognition task, group x drift-rate interaction model  $BF_{10} = 29$ , egocentric eye-gaze task group x drift-rate interaction model  $BF_{10} = 31$ ).

In participants with an IQ above 85, we either found no, or inconclusive evidence, that drift-rate predicted ADHD symptoms on either task (main effect model  $BF_{10}$  for inclusion, face recognition = 0.70, egocentric eye-gaze = 0.25), and no, or inconclusive, evidence for a moderation by group either (interaction model  $BF_{10}$  for inclusion = 1.26, egocentric eye-gaze = 0.42). However, because most of these Bayes factors fall in the range where there is no evidence to support the null or the alternative hypothesis, we cannot know the extent to which the reduced sample size in this cohort contributed to the results.

### *Cross-task correlations*

We were also interested in the extent to which the DDM parameters represent stable traits that are consistent across tasks. To explore this, we compared individual differences in parameters across tasks using Bayesian Pearson correlations (Figure 8). We set a beta prior width of 1 and assumed no hypothesis for the correlation direction. Data from the ASD and non-ASD groups were pooled because Fisher  $r$  to  $z$  transformations suggested no evidence of a difference between the groups on these correlations. There was a clear positive correlation between drift-rates for the two tasks ( $R = 0.5$ ,  $BF_{10} = 10^6$ ). Likewise, we found strong evidence for a positive correlation between boundary separation on the two tasks ( $R = 0.32$ ,  $BF_{10} = 55$ ). We found inconclusive evidence that non-decision time correlated across tasks ( $R = 0.14$ ,  $BF_{10} = 0.36$ ). This pattern of correlations remained when only data from participants with IQs above 85 were analysed (drift-rate,  $R=0.39$ ,  $BF_{10} = 92$ ; boundary separation,  $R = 0.29$ ,  $BF_{10} = 4.8$ ; NDT,  $R= 0.22$ ,  $BF_{10} = 0.93$ ). To summarise, individual differences in drift-rate (processing efficiency) and boundary separation (caution) were correlated across tasks, while non-decision times (perceptual encoding/motor output) were either more task specific or noisier.

## Discussion

Accumulator models, such as the drift-diffusion model, are gaining popularity in clinical research due to their potential to isolate mechanisms underlying the decision-making process. Our interest here was whether autistic individuals display global and stable differences in DDM parameters such as drift-rate (processing efficiency), boundary separation (caution) and non-decision time (perceptual encoding, motor output). Using data from two face processing tasks, we measured differences in model parameters between autistic and non-autistic adolescents across a wide range of IQ and in a sub-group with IQ above 85. In the group with IQ above 85, we found that autistic participants exhibited lower drift-rates and narrower boundary separation than IQ-matched non-autistic adolescents, on both tasks. This suggests that autistic individuals did not process information as efficiently, and were less cautious, than non-autistic individuals. We did not find any group differences in the time associated with perceptual encoding and motor output (non-decision time). When participants with low IQ were also included in the analysis, the results were not as consistent across tasks, and we found evidence of boundary difference in one task and drift-rates differences in the other. There were still no differences in non-decision time.

Crucially, if we had only analysed the means of RT and accuracy our interpretation of the results would have been quite different. For participants with an IQ above 85, we found differences in accuracy but not in RT, which could be interpreted as similar processing speed between the groups but that the autistic participants made more errors. The DDM provided a different interpretation – that processing speed *did* differ between the groups, but so did caution, which could have masked task-related differences in raw RT. Furthermore, using a classification procedure we found that DDM parameters were better able to differentiate between the groups on the two face-processing tasks than standard behavioural measures. The aim of the classification was not to provide a diagnostic measure of ASD, but merely to demonstrate that model parameters may offer greater discriminability relative to traditional behaviour measures, at least in the two tasks examined in this study.

### *Limitations*

Before discussing the results further, it is important to note that the data were not collected with modelling in mind and therefore we were limited in our possible analyses. In particular, we had to pool trials across conditions so that the parameter recovery would be reliable. This was the main study limitation, which we discuss in detail below. This is likely to be a frequent problem for post-hoc data modelling in any domain, as it is rare to collect an ideal number of trials for modelling,

especially in clinical populations. As modelling grows in accessibility, we hope it might become more common to plan trial numbers with modelling in mind.

*Drift-rate: a measure of neural variability or task related performance?*

One of our main motivations for this study was the previous successful application of DDM to ADHD research. Recent studies have shown that increased intra-participant variability in ADHD can be explained by a reduction in drift-rate (Karalunas et al., 2012; Karalunas et al., 2013; Metin et al., 2013; Salum et al., 2014). Given reports that autistic individuals may also show increased intra-participant variability (Geurts et al., 2008; Karalunas et al., 2014) and a recent study that found lower drift-rates in autistic children (Karalunas et al., 2018), we anticipated that drift-rates may also be reduced in our autistic participants. In line with this prediction, drift-rates were reduced for both tasks in our ASD group with IQ above 85, which was most comparable to the participant sample used in previous studies. However, we also found that autistic participants had narrower boundary separation than the non-autistic group. This is contrary to the only other two studies that have applied the DDM to ASD data, where *wider* boundary separation was found in the ASD group (Karalunas et al., 2018; Pirrone et al., 2016). It should be noted that both of these studies used a different autistic population (intellectually able adults and children) to ours (adolescents with a wide range of intellectual ability).

It is notable that Pirrone et al. used a low level perceptual discrimination task (orientation discrimination), which has previously been found to be enhanced in ASD (Dickinson, Bruyns-Haylett, Smith, Jones, & Milne, 2016). In contrast, we used two socially-relevant tasks that centred on faces, and reflect substantial literatures finding impairment in both face recognition (Weigelt, Koldewyn, & Kanwisher, 2012) and processing eye gaze (Nation & Penny, 2008). The reduction in drift-rate (information processing efficiency) that we observed in our autistic participants could be specific to face processing. Further, autistic participants may have been less cautious in responding in these particular tasks because they found the face stimuli more aversive (e.g. Tottenham et al., 2013). Indeed, this might explain why our results for boundary separation were different from those of Pirrone et al and Karalunas et al., both of whom used non-social stimuli. One of our main areas of interest is whether DDM parameters measured for a particular task or group would generalise to other tasks and other sub-populations of that group. Unfortunately, we did not have access to a non-face processing task that was also suitable for DDM fitting to act as a control. However, the disparities between the Pirrone et al. (2016) and Karalunas et al. (2018) studies and ours could suggest that parameters depend on task domain. Future research should explore this further with a wider variety of tasks. It is worth noting, however, that for the two face-processing tasks we

analysed, the boundary separation and drift-rate parameters were reasonably well correlated across individuals. It is also the case that even if the results do not generalise to non-socially related tasks, the DDM findings still challenged the traditional interpretation that any differences in RT/Accuracy are solely due to task related processing.

Increased intra-participant variability in ADHD has been found in a number of tasks and so may represent a global deficit in information processing that is related to lapses in attention (Karalunas et al., 2014). Differences in drift-rates in ASD have currently been explored on fewer tasks, although it is likely that increased variability in the RT distribution, in absence of increased accuracy, will often translate to lower drift-rate under the DDM. It has previously been suggested that increased intra-participant variability in ASD may be driven by a sub-set of autistic participants who have elevated ADHD traits (Karalunas et al., 2014). However, in our autistic sample drift-rate did not correlate with a parent reported measure of hyperactivity, inattention and impulsivity, suggesting that the differences we observed between the ASD and non ASD groups were not being driven by co-occurring ADHD traits. This finding is also supported by previous work by Karalunas et al (2018) who also found that differences in drift-rate in autistic participants were not explained by ADHD symptoms.

Processing efficiency, which is reflected in differences in drift-rate, can be due to a number of factors. For individuals with ADHD, it is generally thought to represent a global difficulty in information processing, in particular an increase in attentional lapses (Karalunas et al., 2014). At a neural level, this could correspond to increases in neural noise and differences in oscillatory brain rhythms (Castellanos et al., 2005; Di Martino et al., 2008; Tamm et al., 2012). However, in our study we used measures of face processing, an area often associated with difficulties in autistic populations. Thus, decreased drift-rates might be more likely to correspond to difficulties in task-specific processing rather than global attentional problems. This may also explain why for non-autistic adolescents, IQ was strongly related to task performance (drift-rate), but this relationship was not present for autistic adolescents where task performance may have been influenced by factors beyond general intellectual ability. As noted earlier, future research exploring a wider range of tasks will help to disentangle these two hypotheses. It is worth considering for future research that a decrease in drift-rate can either be due to deficits in task performance or due to a global processing defects related to increased variability, which could potentially be seen as a limitation of the modelling approach. Of course, this limitation would also apply to traditional behavioural measures, because increased variability in RT distributions tend to result in average RTs that are longer.

### *Estimation of ADHD symptoms*

An important limitation of our study is that the ADHD symptom measure we used from the SDQ did not allow us to explore the separate effects of inattention and hyperactivity, the former of which is most commonly associated with cognitive performance deficits (Castellanos, Sonuga-Barke, Milham, & Tannock, 2006). However, the SDQ is well validated and the items that load on the ADHD factor (inattention, hyperactivity, impulsivity) show good internal consistency across individuals (Goodman, 2001). Future research is needed to examine whether individual differences in hyperactivity, impulsivity and inattention show different interactions with DDM parameters in autistic individuals. It might be predicted that inattention is more strongly related to drift-rate, while impulsivity is related to caution. Unfortunately, such a fine-grained analysis was beyond the scope of the current study.

### *Effects of IQ variability and participant heterogeneity*

Our study had a number of advantages, including a reasonably large sample size and a wide range of intellectual ability. The majority of research looking at cognitive performance in ASD tends to focus on individuals with average to high IQ, which is not representative of the autistic population as a whole (Brugha et al., 2018). In our study we had wide range of ability in the ASD group and an IQ-matched comparison group. However, many of the individuals with lower IQ had additional clinical diagnoses, as can often be the case with individuals with low IQ. Therefore, in order to ensure that our results were not being driven by individuals with additional educational needs or clinical diagnoses, we repeated all of the analyses for a restricted range of participants with IQ above 85.

Mostly, the results remained consistent with those collected from the whole cohort, although differences in drift-rate and boundary were more evident in the group with IQ above 85, and we also found differences in raw accuracy. This suggests that group differences were generally larger in the subgroup with higher IQ. It is possible that because our comparison group was quite heterogeneous, this masked some of the differences between the groups. The heterogeneity of our comparison group meant that we did not have the power to look at different subgroups within the lower IQ range, for example, those with different types of learning disability, but it is an interesting question for future research. It is also possible that the quality of the data was not as good for individuals with low IQ and this meant it was more difficult to detect group differences.

Another interesting question for future research is whether there are any sex differences in the parameter differences found for the ASD group. Due to the small number of female participants we were not able to explore this question in the present study, but given increasing evidence that

the female phenotype in autism might differ from the male (e.g. Mandy et al., 2012), it is certainly an important factor to consider in the future.

#### *Effects of collapsing over conditions on model parameters*

An important limitation of the current study was that there were relatively few trials per participant to use for model fitting. This meant that we had to collapse across experimental conditions in order to maximise data points, which was not ideal. Collapsing across conditions would not be expected to affect our treatment of the boundary separation and non-decision time parameters, as these are commonly fixed across conditions when trials are presented in an unpredictable order. It is traditionally assumed that the boundary is set before the trial begins and if the conditions are inter-mixed (as ours mostly were) then the participant would not have an opportunity to adjust their boundary based on condition. There are alternative models that assume dynamic boundaries within a trial and across a block of trials, however, it is rare for these models to improve data fits and so it remains more common practice to assume a fixed boundary (Voskuilen, Ratcliff, & Smith, 2016).

For drift rates, we would expect differences between easier and more difficult conditions. The differences in drift-rate we observed may well have been driven by a subset of conditions, which is an important limitation of the current study. Potentially, different drift rate fits for each condition could also have had consequences for how other parameters were fitted. In addition to low trial numbers, our reasoning for collapsing across conditions was that we were particularly interested in any global and stable differences in processing. This would be predicted by certain neurocognitive theories of ASD, for example, the idea that autistic individuals have increased neural noise (Milne, 2011; Simmons et al., 2007). Under these theories, a reduction in drift-rate should occur no matter what task or condition. In terms of the results of the current study, we can conclude that there were differences in drift-rate between the groups and that the model parameters were better at detecting these differences than the behavioural measures. Future research is required to explore whether drift-rate differences in ASD were observed are condition-specific or task-specific. This would be particularly informative if the manipulation of conditions was consistent across tasks (e.g. comparing social and non-social conditions in each task).

#### *Trial numbers for reliable estimation of cognitive measures*

One general limitation of fitting accumulator models is that the trial numbers required to obtain a reliable estimate of model parameters are not always feasible in clinical research. Bayesian hierarchical fitting methods, like the HDDM, go some way in easing this problem (Wiecki et al., 2013). It is important to note that low trial numbers are a general issue even if standard measures of

mean accuracy and RT are used. Research from our lab has shown that many more trial numbers are required to obtain reliable estimates of performance from classic forced-choice tasks than are often used in the literature (Hedge, Powell, & Sumner, 2018). For example, in the flanker task, which is a common measure of response inhibition, around 150 trials are required to obtain a stable flanker effect for each individual. However, even with 150 trials, the reliability of the flanker effect across sessions is low (intra-class correlation = 0.5). Thus, the conflicting findings in the ASD literature for forced-choice perceptual and cognitive tasks, which are commonly ascribed to the heterogeneity of the disorder and low sample sizes, could very well reflect unreliable measures and measurements. That is, if the same group of participants were tested a week later different results might be obtained and different conclusions drawn. It remains to be seen whether DDM parameters could help to increase reliability when trial numbers are relatively low.

#### *Potential advantages of accumulator models for clinical research*

This is the first study apply the DDM to Autism research using social stimuli and we found that our interpretation of the results would have been quite different if we had only analysed the traditional behavioural measures. We also found that model parameters were fairly consistent across individuals and tasks. Parameters derived from cognitive modelling may have an advantage over measures of RT and accuracy if they tap into more stable and reliable traits across individuals. Certainly in the ASD literature, where conflicting experimental findings are common, any measure with the potential to increase reliability and discriminability is worth pursuing. Further, given that many task-related differences between autistic and non -autistic groups reduce with age (e.g. Happe, 1995), examining drift-rate could increase sensitivity to detect these differences later in development. It is important to note that these models cannot add 'extra' variance (or 'signal') that is not present in the original accuracy and RT data. What they can do is structure and organise this variance into more psychologically meaningful terms and provide a principled way to utilise data from across the whole RT distribution. Isolating more psychologically meaningful factors from RT and accuracy data may help to identify key areas of difficulty across conditions and individuals. This in turn may lead to the development of more effective and individualised cognitive training paradigms, which at present mainly focus on measuring improvements in mean accuracy and RT (e.g. Powell, Wass, Erichsen, & Leekam, 2016).

## References

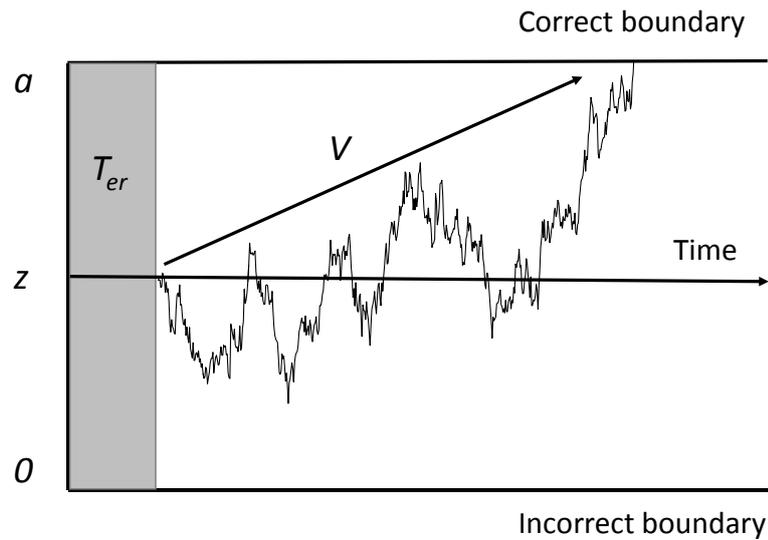
- Abikoff, H. (1991). Cognitive training in ADHD children: Less to it than meets the eye. *Journal of Learning Disabilities, 24*(4), 205-209.
- American Psychiatric Association. (2013). *DSM 5* (5th ed.). Washington, DC: American Psychiatric Association.
- Ashwin, C., Ricciardelli, P., & Baron-Cohen, S. (2009). Positive and negative gaze perception in autism spectrum conditions. *Social Neuroscience, 4*(2), 153-164.
- Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *The Lancet, 368*(9531), 210-215.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113*(4), 700.
- Brugha, T. S., Spiers, N., Bankart, J., Cooper, S.-A., McManus, S., Scott, F. J., . . . Tyrer, F. (2018). Epidemiology of autism in adults across age groups and ability levels. *British Journal of Psychiatry, 209*(6), 498-503. doi:10.1192/bjp.bp.115.174649
- Castellanos, F. X., Sonuga-Barke, E. J., Milham, M. P., & Tannock, R. (2006). Characterizing cognition in ADHD: beyond executive dysfunction. *Trends in Cognitive Sciences, 10*(3), 117-123.
- Castellanos, F. X., Sonuga-Barke, E. J. S., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of Attention-Deficit/Hyperactivity Disorder-Related Intra-Individual Variability. *Biological psychiatry, 57*(11), 1416-1423. doi:10.1016/j.biopsych.2004.12.005
- Charman, T., Jones, C. R., Pickles, A., Simonoff, E., Baird, G., & Happé, F. (2011). Defining the cognitive phenotype of autism. *Brain research, 1380*, 10-21.
- Di Martino, A., Ghaffari, M., Curchack, J., Reiss, P., Hyde, C., Vannucci, M., . . . Castellanos, F. X. (2008). Decomposing Intra-Subject Variability in Children with Attention-Deficit/Hyperactivity Disorder. *Biological psychiatry, 64*(7), 607-614. doi:10.1016/j.biopsych.2008.03.008
- Dickinson, A., Bruyins-Haylett, M., Smith, R., Jones, M., & Milne, E. (2016). Superior orientation discrimination and increased peak gamma frequency in autism spectrum conditions. *Journal of abnormal psychology, 125*(3), 412.
- Elgar, K., Campbell, R., & Skuse, D. (2002). Are you looking at me? Accuracy in processing line-of-sight in Turner syndrome. *Proceedings of the Royal Society of London B: Biological Sciences, 269*(1508), 2415-2422.
- Epstein, J. N., Langberg, J. M., Rosen, P. J., Graham, A., Narad, M. E., Antonini, T. N., . . . Altaye, M. (2011). Evidence for higher reaction time variability for children with ADHD on a range of cognitive tasks including reward and event rate manipulations. *Neuropsychology, 25*(4), 427.
- Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1): Wiley, New York.
- Forgeot d'Arc, B., Delorme, R., Zalla, T., Lefebvre, A., Amsellem, F., Moukawane, S., . . . Ramus, F. (2017). Gaze direction detection in autism spectrum disorder. *Autism, 21*(1), 100-107.
- Forstmann, B., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Psychology, 67*(1), 641.
- Gargaro, B. A., Rinehart, N. J., Bradshaw, J. L., Tonge, B. J., & Sheppard, D. M. (2011). Autism and ADHD: how far have we come in the comorbidity debate? *Neuroscience & Biobehavioral Reviews, 35*(5), 1081-1088.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science, 7*(4), 457-472.
- Geurts, H. M., Grasman, R. P., Verté, S., Oosterlaan, J., Roeyers, H., van Kammen, S. M., & Sergeant, J. A. (2008). Intra-individual variability in ADHD, autism spectrum disorders and Tourette's syndrome. *Neuropsychologia, 46*(13), 3030-3041.

- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581-586.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345.
- Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3), 839-843. doi:10.1148/radiology.148.3.6878708
- Happe, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child development*, 66(3), 843-855.
- Happe, F. G., & Ronald, A. (2008). The 'fractionable autism triad': a review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychol Rev*, 18(4), 287-304. doi:10.1007/s11065-008-9076-8
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav Res Methods*, 50(3), 1166-1186.
- JASP-Team. (2016). JASP Version 0.8.0.0 Windows 7.
- Jeffreys, H. (1961). *Theory of probability*. New York: Oxford University Press.
- Johnson, K. A., Kelly, S. P., Bellgrove, M. A., Barry, E., Cox, M., Gill, M., & Robertson, I. H. (2007). Response variability in attention deficit hyperactivity disorder: evidence for neuropsychological heterogeneity. *Neuropsychologia*, 45(4), 630-638.
- Joseph, R. M., & Tanaka, J. (2003). Holistic and part-based face recognition in children with autism. *Journal of Child Psychology and Psychiatry*, 44(4), 529-542.
- Karalunas, S. L., Geurts, H. M., Konrad, K., Bender, S., & Nigg, J. T. (2014). Annual research review: Reaction time variability in ADHD and autism spectrum disorders: measurement and mechanisms of a proposed trans-diagnostic phenotype. *Journal of Child Psychology and Psychiatry*, 55(6), 685-710. doi:10.1111/jcpp.12217
- Karalunas, S. L., Hawkey, E., Gustafsson, H., Miller, M., Langhorst, M., Cordova, M., . . . Nigg, J. T. (2018). Overlapping and Distinct Cognitive Impairments in Attention-Deficit/Hyperactivity and Autism Spectrum Disorder without Intellectual Disability. *Journal of abnormal child psychology*, 1-12.
- Karalunas, S. L., Huang-Pollock, C. L., & Nigg, J. T. (2012). Decomposing attention-deficit/hyperactivity disorder (ADHD)-related effects in response speed and variability. *Neuropsychology*, 26(6), 684.
- Karalunas, S. L., Huang-Pollock, C. L., & Nigg, J. T. (2013). Is reaction time variability in ADHD mainly at low frequencies? *Journal of Child Psychology and Psychiatry*, 54(5), 536-544.
- Keen, D., & Ward, S. (2004). Autistic spectrum disorder a child population profile. *Autism*, 8(1), 39-48.
- Klein, C., Wendling, K., Huettner, P., Ruder, H., & Peper, M. (2006). Intra-subject variability in attention-deficit hyperactivity disorder. *Biological psychiatry*, 60(10), 1088-1097.
- Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: a meta-analytic review of 319 studies. *Clinical psychology review*, 33(6), 795-811.
- Kuiper, M. W. M., Verhoeven, E. W. M., & Geurts, H. M. (2016). The role of interstimulus interval and "Stimulus-type" in prepotent response inhibition abilities in people with ASD: A quantitative and qualitative review. *Autism Research*, 9(11), 1124-1141. doi:10.1002/aur.1631
- Kuntsi, J., Oosterlaan, J., & Stevenson, J. (2001). Psychological mechanisms in hyperactivity: I response inhibition deficit, working memory impairment, delay aversion, or something else? *Journal of Child Psychology and Psychiatry*, 42(02), 199-210.
- Langdell, T. (1978). Recognition of faces: An approach to the study of autism. *Journal of Child Psychology and Psychiatry*, 19(3), 255-268.

- Leekam, S. (2016). Social cognitive impairment and autism: what are we trying to explain? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1686). doi:10.1098/rstb.2015.0082
- Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British journal of developmental psychology*, 15(1), 77-95.
- Leth-Steensen, C., Elbaz, Z. K., & Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychologica*, 104(2), 167-190.
- Leyfer, O. T., Folstein, S. E., Bacalman, S., Davis, N. O., Dinh, E., Morgan, J., . . . Lainhart, J. E. (2006). Comorbid psychiatric disorders in children with autism: Interview development and rates of disorders. *Journal of Autism and Developmental Disorders*, 36(7), 849-861.
- Lord, C., Risi, S., Lambrecht, L., Cook Jr, E. H., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205-223.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659-685.
- Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2012). Sex Differences in Autism Spectrum Disorder: Evidence from a Large Sample of Children and Adolescents. *Journal of Autism and Developmental Disorders*, 42(7), 1304-1313. doi:10.1007/s10803-011-1356-0
- Mayes, S. D., Calhoun, S. L., Mayes, R. D., & Molitoris, S. (2012). Autism and ADHD: Overlapping and discriminating symptoms. *Research in Autism Spectrum Disorders*, 6(1), 277-285.
- Metin, B., Roeyers, H., Wiersema, J. R., van der Meere, J. J., Thompson, M., & Sonuga-Barke, E. (2013). ADHD performance reflects inefficient but not impulsive information processing: A diffusion model analysis. *Neuropsychology*, 27(2), 193.
- Milne, E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked EEG. *Frontiers in psychology*, 2, 51.
- Moustafa, A. A., Kéri, S., Somlai, Z., Balsdon, T., Frydecka, D., Misiak, B., & White, C. (2015). Drift diffusion model of reward and punishment learning in schizophrenia: modeling and experimental data. *Behavioural brain research*, 291, 147-154.
- Mullins, C., Bellgrove, M. A., Gill, M., & Robertson, I. H. (2005). Variability in time reproduction: difference in ADHD combined and inattentive subtypes. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(2), 169-176.
- Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: is it normal? Is it automatic? Is it social? *Development and psychopathology*, 20(01), 79-97.
- Pellicano, E., & Burr, D. C. (2012). When the world becomes too real: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16(10), 504-510.
- Pirrone, A., Dickinson, A., Gomez, R., Stafford, T., & Milne, E. (2016). Understanding Perceptual Judgment in Autism Spectrum Disorder Using the Drift Diffusion Model. *Neuropsychology*.
- Powell, G., Wass, S. V., Erichsen, J. T., & Leekam, S. R. (2016). First evidence of the feasibility of gaze-contingent attention training for school children with autism. *Autism*, 1362361315617880.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59.
- Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, 2(4), 237-279.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347-356.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, 20(4), 260-281.

- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and aging, 19*(2), 278.
- Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*(4), 408-424.
- Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140*(3), 464.
- Rommelse, N. N., Franke, B., Geurts, H. M., Hartman, C. A., & Buitelaar, J. K. (2010). Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *European child & adolescent psychiatry, 19*(3), 281-295.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356-374.
- Russell, V. A., Oades, R. D., Tannock, R., Killeen, P. R., Auerbach, J. G., Johansen, E. B., & Sagvolden, T. (2006). Response variability in attention-deficit/hyperactivity disorder: a neuronal and glial energetics hypothesis. *Behavioral and Brain Functions, 2*(1), 1.
- Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*: Western Psychological Services.
- Salum, G., Sergeant, J., Sonuga-Barke, E., Vandekerckhove, J., Gadelha, A., Pan, P., . . . do Rosario, M. (2014). Specificity of basic information processing and inhibitory control in attention deficit hyperactivity disorder. *Psychological Medicine, 44*(03), 617-631.
- Simmons, D., McKay, L., McAleer, P., Toal, E., Robertson, A., & Pollick, F. (2007). Neural noise and autism spectrum disorders. *Perception, 36*, 119-120.
- Simmons, D., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research, 49*(22), 2705-2739.
- Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(8), 921-929.
- Tamm, L., Narad, M. E., Antonini, T. N., O'Brien, K. M., Hawk, L. W., & Epstein, J. N. (2012). Reaction time variability in ADHD: a review. *Neurotherapeutics, 9*(3), 500-508.
- Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review, 120*(1), 1.
- Tottenham, N., Hertzog, M. E., Gillespie-Lynch, K., Gilhooly, T., Millner, A. J., & Casey, B. J. (2013). Elevated amygdala response to faces and gaze aversion in autism spectrum disorder. *Soc Cogn Affect Neurosci*. doi:10.1093/scan/nst050
- van Ravenzwaaij, D., Brown, S., & Wagenmakers, E.-J. (2011). An integrated perspective on the relation between response speed and intelligence. *Cognition, 119*(3), 381-393.
- Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology, 73*, 59-79. doi:10.1016/j.jmp.2016.04.008
- Wald, A. (1947). Sequential analysis. 1947. *Zbl0029, 15805*.
- Webster, S., & Potter, D. D. (2008). Brief report: Eye direction detection improves with development in autism. *Journal of Autism and Developmental Disorders, 38*(6), 1184-1186.
- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*: Psychological Corporation.
- Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neuroscience & Biobehavioral Reviews, 36*(3), 1060-1084.
- Wiecki, T., Sofer, I., & Frank, M. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics, 7*(14). doi:10.3389/fninf.2013.00014

Zhang, J., Rittman, T., Nombela, C., Fois, A., Coyle-Gilchrist, I., Barker, R. A., . . . Rowe, J. B. (2016). Different decision deficits impair response inhibition in progressive supranuclear palsy and Parkinson's disease. *Brain*, *139*(1), 161-173.



*Figure 1. Schematic representation of the drift-diffusion model.* Information accumulates stochastically with drift-rate ( $v$ ) until it reaches one of two response boundaries, at which point a response is triggered. The accuracy of the response is therefore determined by which bound was reached. Boundary separation is represented by  $a$ . Bias in starting point towards either boundary is given by  $z$ . Non-decision time ( $T_{er}$ ), represents the time taken before and after the accumulation process to perceptually encode the stimulus and to generate a motor response. Reaction time is determined by the time taken to reach a bound plus non-decision time. Overall, higher drift-rates lead to fast and accurate responses, while lower drift rates result in slow and inaccurate responses. Large boundary separations lead to slow and accurate responses and small boundary separations result in fast but inaccurate responses.

	<b>Autistic group (n = 75)</b>	<b>Not Autistic group (n = 46)</b>	<b>Autistic subgroup IQ &lt;85 (n= 51)</b>	<b>Not Autistic subgroup IQ &lt;85 (n= 31)</b>
Age	15.5 ± 0.54	15.5 ± 1.3	15.5 ± 6.1 (15.1 – 16.9)	15.5 ± 7.2 (14.1 – 16.8)
Gender	67 male	45 male	45 male	30 male
Full-scale IQ (WASI)	85 ± 16.6	90 ± 18.9	89.7 ± 11.7 (85-119)	104 ± 9.2 (85-119)

Table 1. Participant demographics and IQ; mean +/- standard deviation. Statistics shown are means and standard deviations. WASI = Wechsler Abbreviated Scale of Intelligence.

	R scale fixed effects	R scale random effects	R scale covariates
Bayesian ANOVA	0.5	1	0.354
Bayesian Regression	n/a	n/a	0.354
Bayesian Correlation	n/a	n/a	0.354

Table 2. Default JASP priors used for analyses.

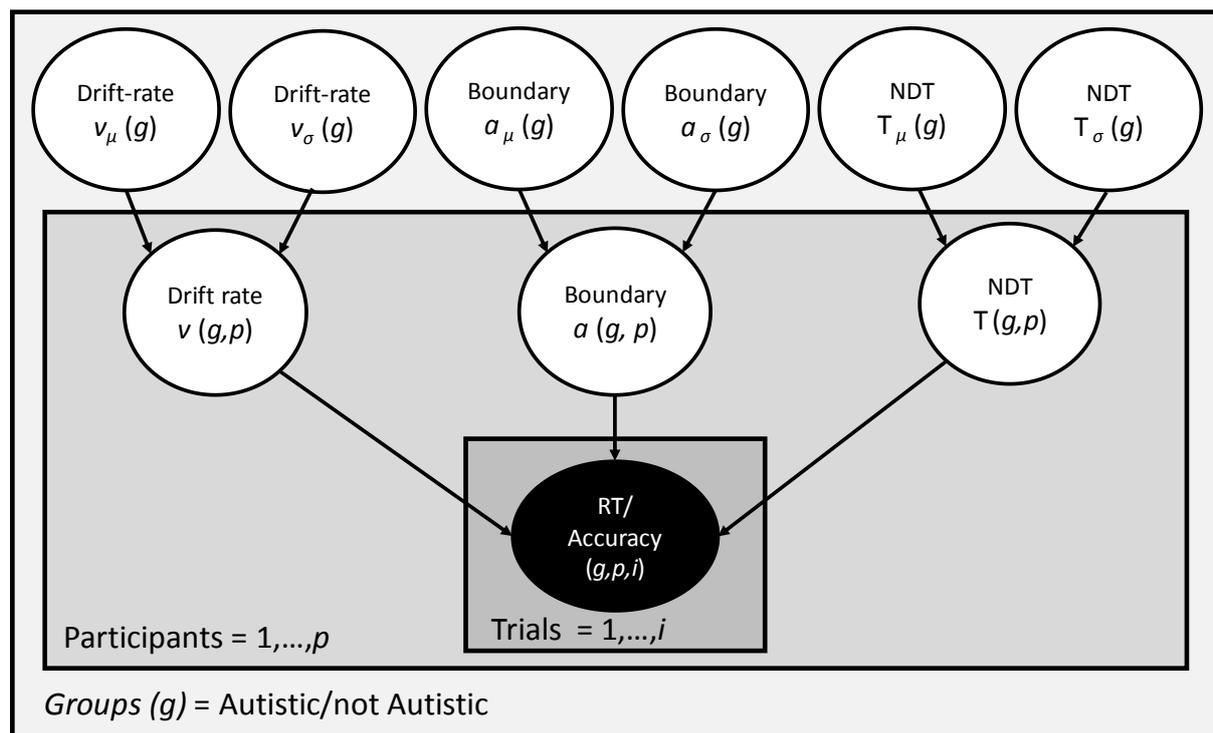
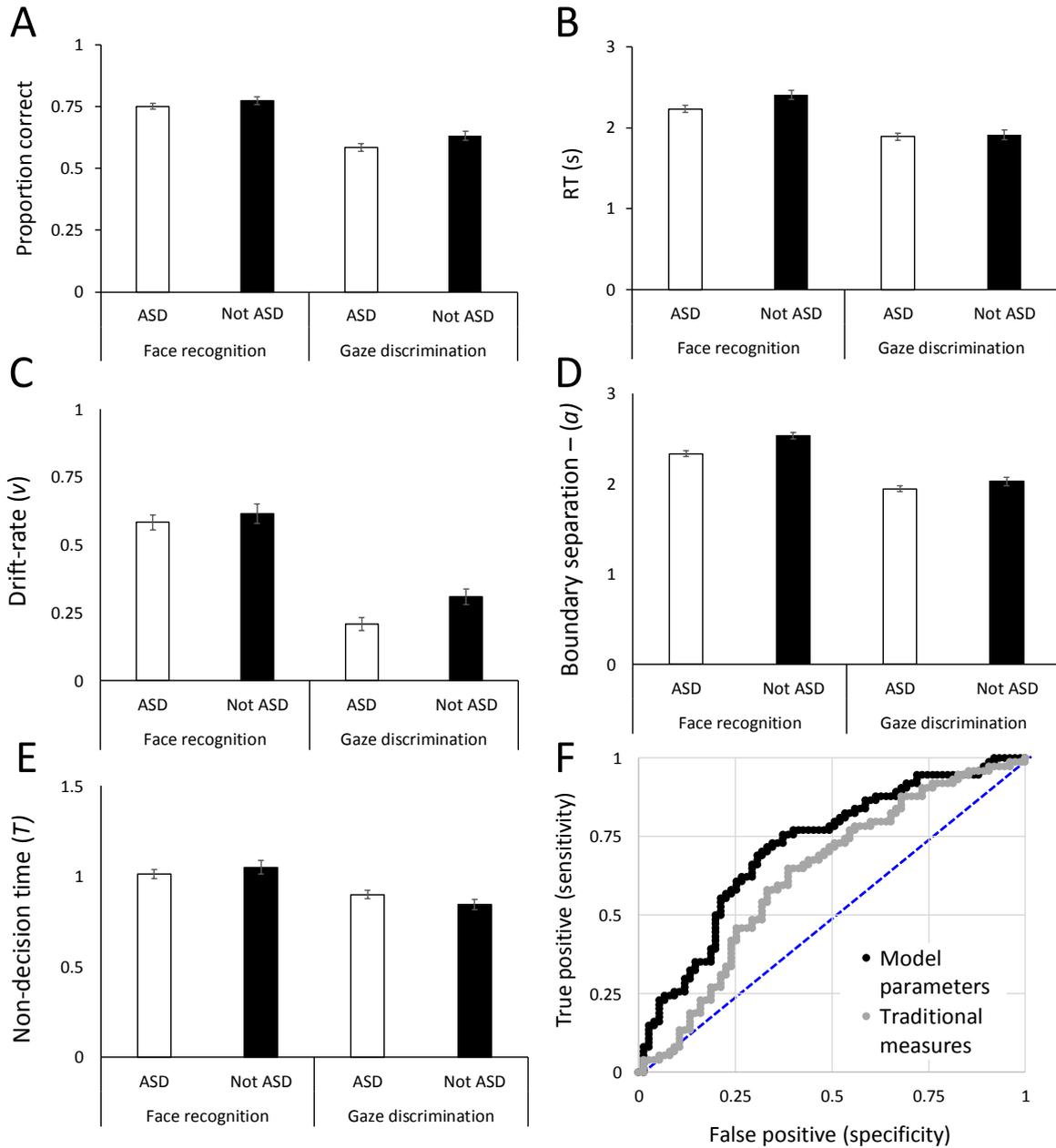
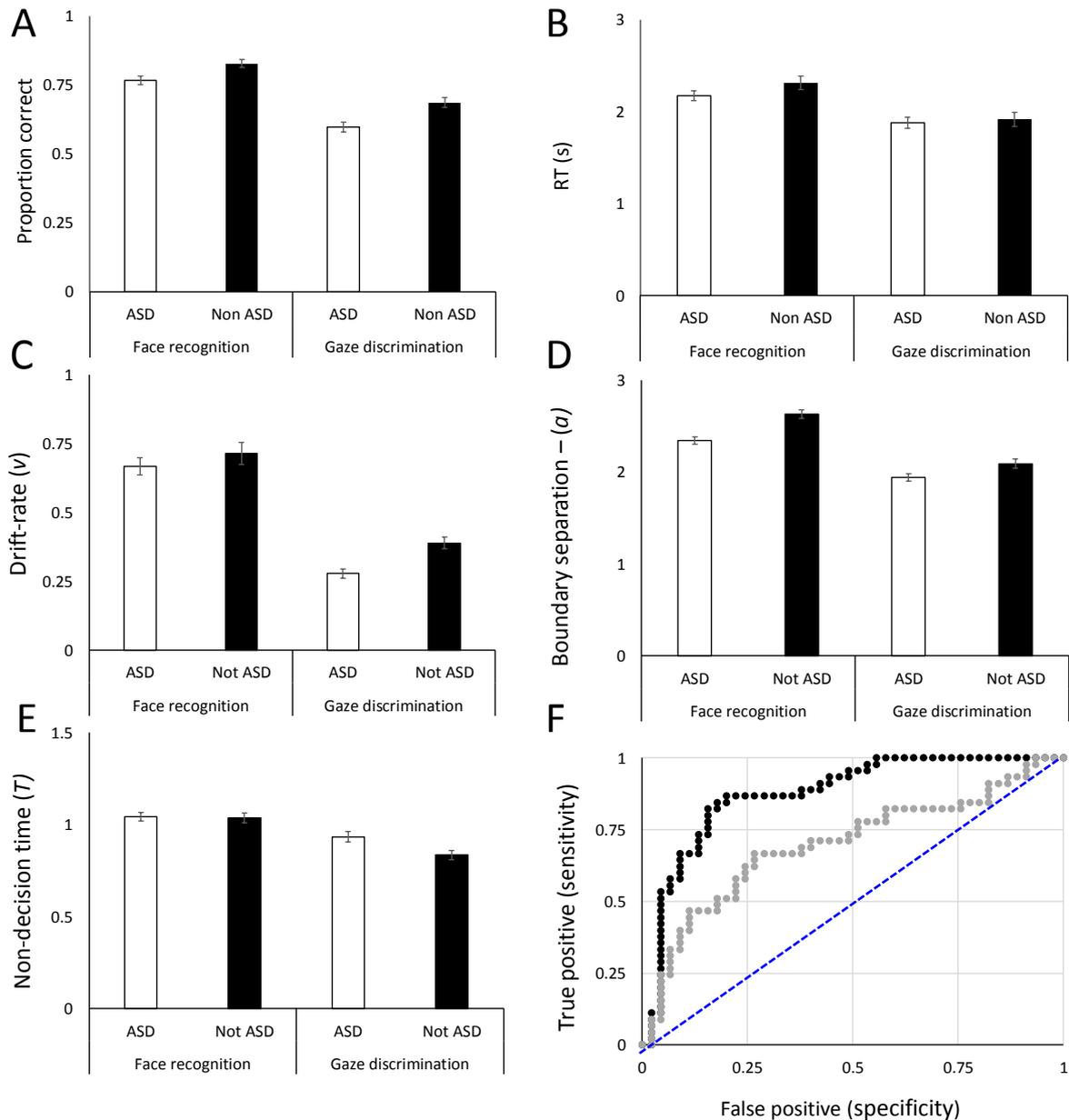


Figure 2. Schematic representation of fitted model. The black node indicates the observed behavioural data (RT/Accuracy) for each group (g), participant (p) and trial (i). Above this are individual level nodes that represent parameter distributions for each participant (p). We fitted three DDM parameters at this level: drift-rate 'v', boundary separation 'a', and non-decision time (NDT) 'T'. Above this are the group level nodes for the same parameters, represented by distributions for each group (ASD, not ASD) that are defined by a mean  $\mu$  and standard deviation  $\sigma$ .



**Figure 3. (A-B) Traditional behavioural measures.** Proportion correct (accuracy) and RT for face recognition and gaze discrimination tasks for Autistic group (white bars) and not Autistic group (black bars), collapsed across conditions. No evidence for main effects or interactions were found (BFs all <3). **(C-E) Estimated DDM parameters across the groups and tasks.** Evidence for group differences were found for drift rate and boundary (see text) **(F) Group classification.** Receiver operating characteristics (ROC) curves for logistic classifier based on model parameters (black line) and traditional behavioural measures (grey lines), for autism classification. Blue line represents chance level performance. For A-E, error bars show standard error.



**Figure 4.** (A-B) Traditional behavioural measures for participants with an IQ above 85. Proportion correct (accuracy) and RT for face recognition and gaze discrimination tasks for Autistic group (white bars) and not Autistic group (black bars), collapsed across conditions. No evidence for main effects or interactions were found (BFs all <3). (C-E) Estimated DDM parameters across the groups and tasks, for participants with an IQ above 85. Evidence for group differences were found for drift rate and boundary (see text) (F) Group classification for participants with an IQ above 85. Receiver operating characteristics (ROC) curves for logistic classifier based on model parameters (black line) and traditional behavioural measures (grey lines), for Autism classification. Blue line represents chance level performance. For A-E, error bars show standard error.

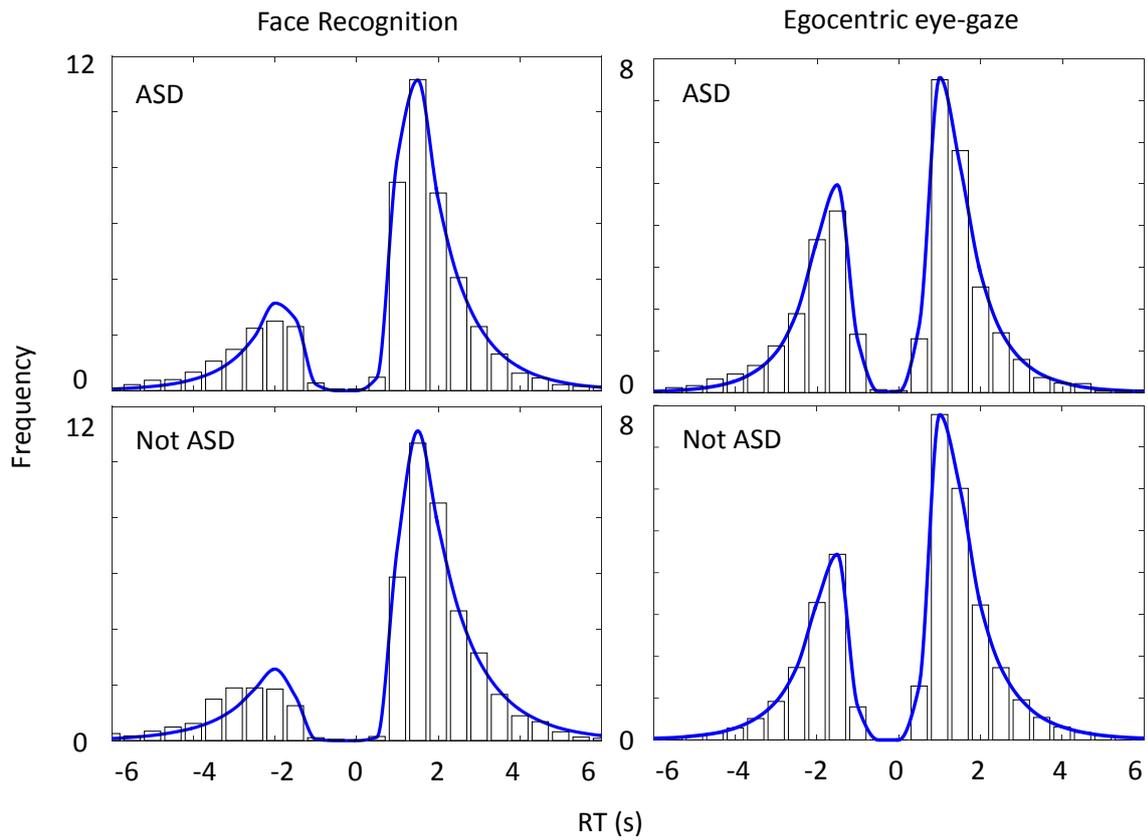


Figure 5. *Model fits - posterior predictive distributions across tasks and groups.* Positive values on the x-axis show correct RTs distributions, negative values show error RT distributions. Observed data from the tasks are shown in the black/white histograms, simulated data based on model fits are represented by blue lines.

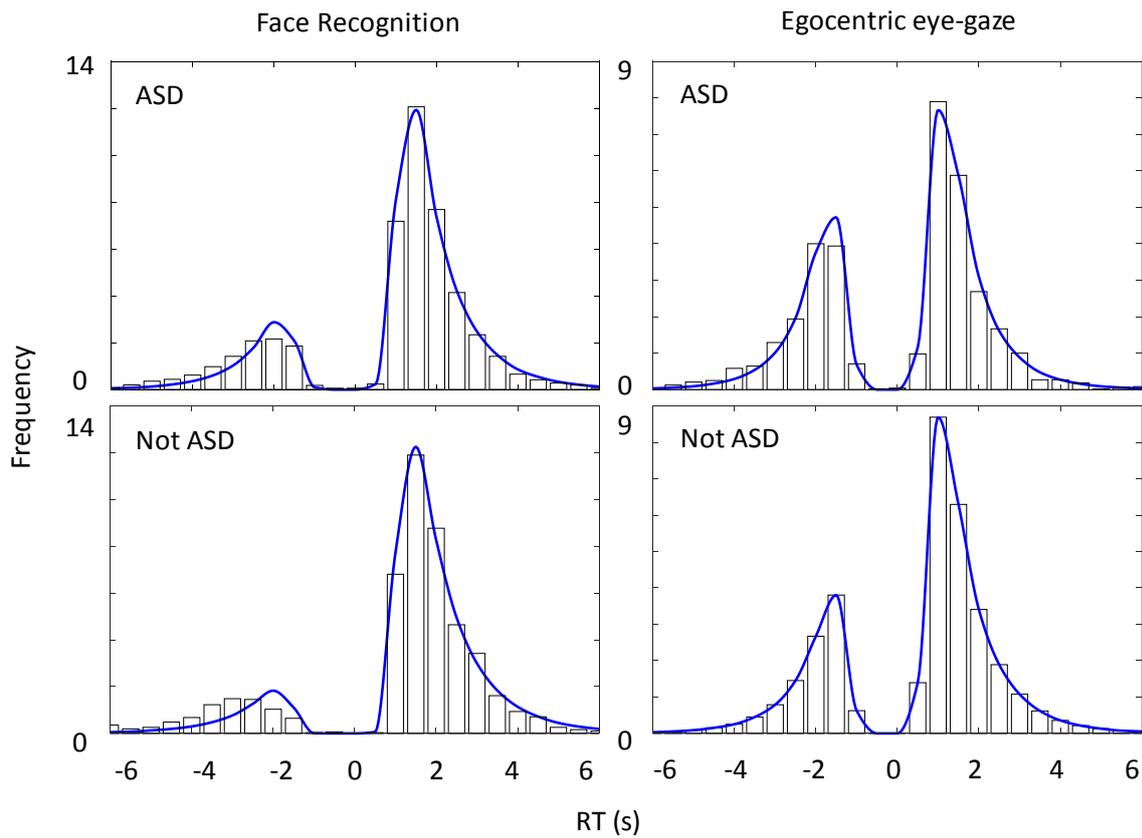


Figure 6. Model fits for participants with IQ above 85. Posterior predictive distribution across tasks and groups. Positive values on the x-axis show correct RTs distributions, negative values show error RT distributions. Observed data from the tasks are shown in the black/white histograms, simulated data based on model fits are represented by blue lines.

	Accuracy	RT 10	RT 30	RT 50	RT 70	RT 90	RT error
Autistic face recognition	0.001	0.008	0.001	0.020	0.037	0.221	0.468
Not Autistic face recognition	0.002	0.007	0.013	0.015	0.038	0.288	0.798
Autistic eye gaze	0.003	0.006	0.013	0.021	0.049	0.401	0.077
Not Autistic eye gaze	0.002	0.005	0.016	0.018	0.056	0.402	0.128

Table 3. *Mean Square Errors (MSEs)* between the observed data and the data simulated from the best fitting model parameters. Column 1 shows mean accuracy, columns 2-6 show correct RT quantiles, and column 7 shows median RT error.

	Accuracy	RT 10	RT 30	RT 50	RT 70	RT 90	RT error
Autistic face recognition	0.002	0.011	0.011	0.016	0.027	0.196	0.568
Not Autistic face recognition	0.002	0.006	0.010	0.012	0.037	0.334	1.081
Autistic eye gaze	0.004	0.004	0.012	0.019	0.054	0.459	0.086
Not Autistic eye gaze	0.003	0.006	0.021	0.015	0.024	0.392	0.148

Table 4. *Mean Square Errors (MSEs)* between the observed data and the data simulated from the best fitting model parameters, for participants with IQ above 85. Column 1 shows mean accuracy, columns 2-6 show correct RT quantiles, and column 7 shows median RT error.

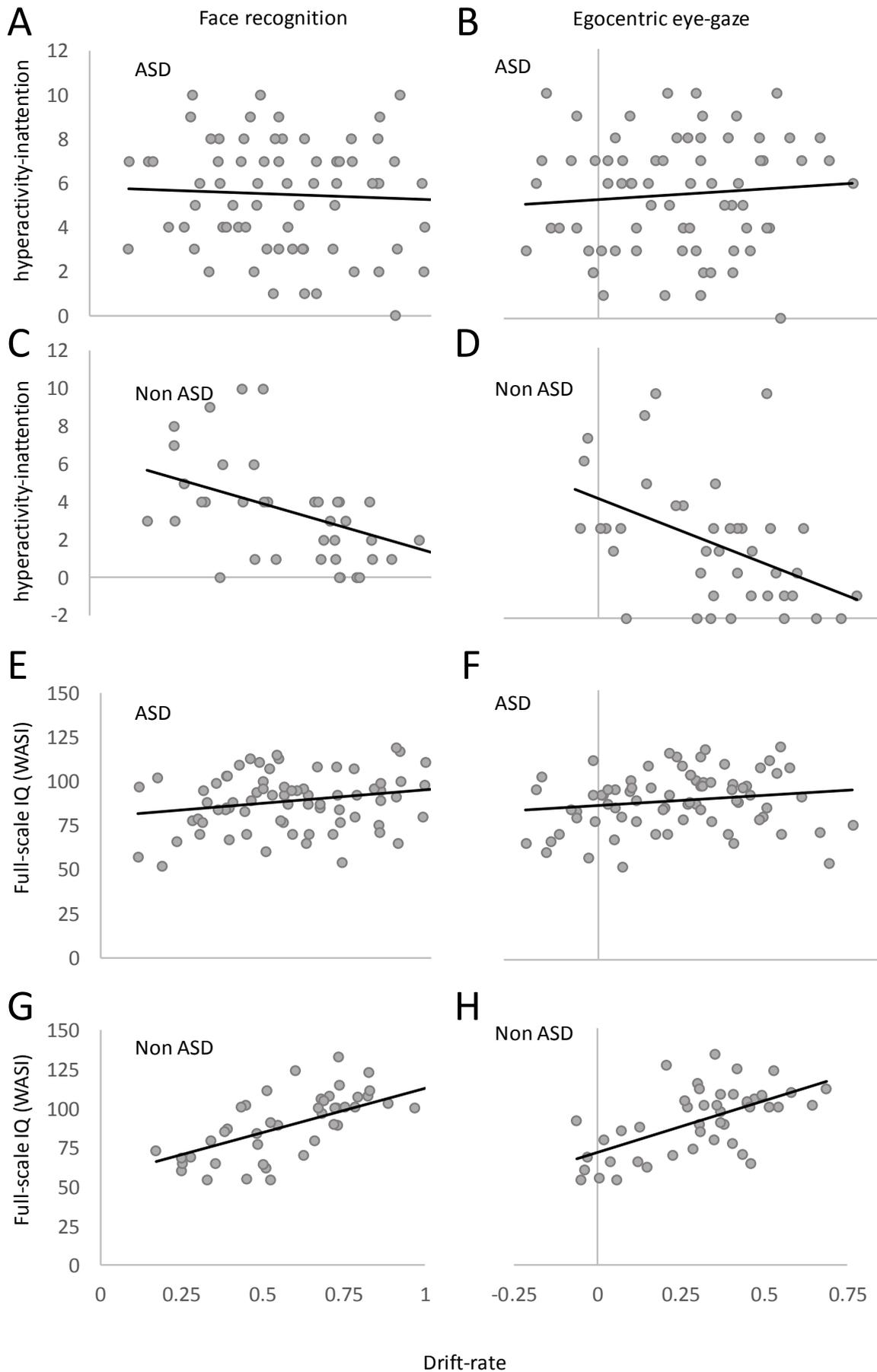


Figure 7. Relationship between hyperactivity-inattention (ADHD symptoms) and drift-rate and full-scale IQ and drift-rate, split by task and group. There was no relationship between hyperactivity-inattention and drift rate in autistic participants (A-B). Whereas, there was a strong negative relationship between hyperactivity-inattention and drift-rate in non-autistic participants (C-D). There was no relationship between drift-rate and IQ in autistic participants (E-F), but a strong relationship between drift-rate and IQ in non-autistic participants (G-H).

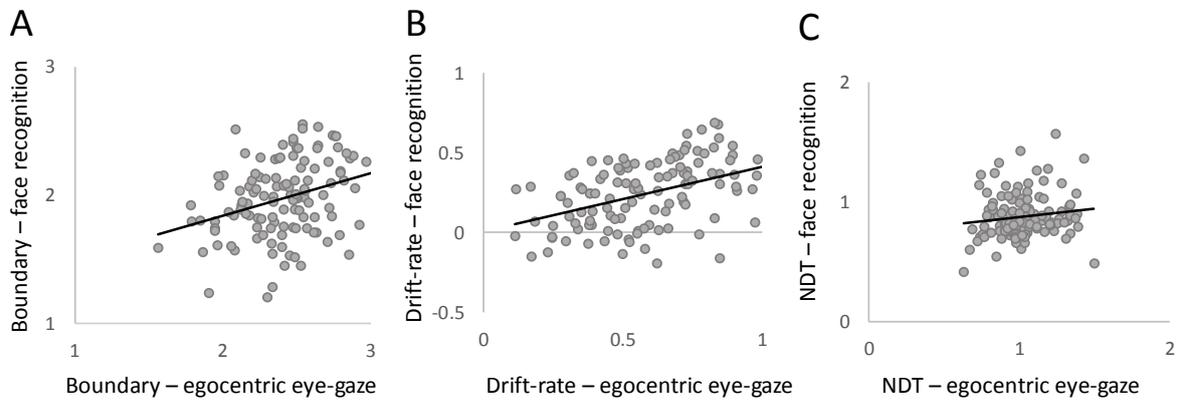


Figure 8. Cross task correlations between estimated DDM parameters. Boundary separation (A) and drift-rate (B) correlated positively across tasks. There was no evidence that non-decision times (C) correlated across tasks.