What can performance in the IEDS task tell us about attention shifting in clinical groups?

James M. Yearsley[a], Sebastian B. Gaigg[a], Dermot M. Bowler[a], Melanie Ring[ab] & Corinna Haenschel[a]

a Department of Psychology,

City, University of London,

EC1V 0HB

b Department of Child and Adolescent Psychiatry,

University Hospital Carl Gustav Carus,

Technische Universität Dresden,

Dresden, Germany

Author Note

Corresponding author: Dr James M Yearsley, Department of Psychology, City, University of London, EC1V 0HB, UK. james.yearsley@city.ac.uk

Data and code connected to this paper can be found on the Open Science Framework https://osf.io/cg2m3/

Abstract

The Intra-Extra-dimensional set shift task (IEDS) is a widely used test of learning and attention, believed to be sensitive to aspects of executive function. The task proceeds through a number of stages, and it is generally claimed that patterns of errors across stages can be used to discriminate between reduced attention switching and more general reductions in rates of learning. A number of papers have used the IEDS task to argue for specific attention shifting difficulties in Autism Spectrum Disorder (ASD) and Schizophrenia, however it remains unclear how well the IEDS really differentiates between reduced attention shifting and other causes of impaired performance. To address this issue, we introduce a simple computational model of performance in the IEDS task, designed to separate the competing effects of attention shifting and general learning rate. We fit the model to data from ASD and comparison individuals matched on age and IQ, as well as to data from four previous studies which used the IEDS task. Model fits do not show consistent evidence for reductions in attention shifting rates in ASD and Schizophrenia. Instead, we find performance is better explained by differences in learning rate, particularly from punishment, which we show correlates with IQ. We therefore argue that the IEDS task is not a good measure of attention shifting in clinical groups.

Lay Summary

The Intra-Extra-Dimensional Set shift task (IEDS) is often given to autistic individuals, who tend to make more errors relative to comparison groups. This higher error rate is taken to mean that autistic individuals struggle with attention control. Our computational model of the IEDS shows that the performance of ASD and some other clinical groups can be explained instead by differences in learning rate, rather than differences in attention control.

Keywords: ASD, Schizophrenia, Extra-Dimensional Shift

What can performance in the IEDS task tell us about attention shifting in clinical groups?

The Intra-Extra-Dimensional Set Shift task (IEDS) is a widely used computer-based test of learning and attention switching, which forms part of the Cambridge Neuropsychological Test Automated Battery (Downes et al, 1989). It has been successfully employed on individuals with frontal lobe lesions (Pantelis et al 1999), Schizophrenia (Jazbec et al 2007; Barch et al, 2009), Parkinson's disease (Downes et al 1989), ASD (Ozonoff et al 2004), and learning impairments (Hughes et al 1993), as well as on non-human animals such as rats (Shypp & Eimas 1964). One advantage of the IEDS task over other measures of attention switching, such as the Wisconsin Card Sorting Task (WCST; Hill, 2004), is that it proceeds through a number of stages of increasing 'complexity' thus allowing researchers to identify the specific stage when performance drops. Of particular interest in this context is the switch from intra-dimensional (ID) stages, which require participants to maintain their attention on a specific set of stimulus features, to extra-dimensional (ED) stages that require a shift of attention to stimulus features that were previously irrelevant. This distinction is thought to allow for a differentiation between impaired attention switching and other learning impairments based on the patterns of errors that participants make at different stages of the task.

A number of studies suggest that individuals with ASD or Schizophrenia have difficulties at the ED stage of the task (Hughes et al, 1994; Jazbec et al 2007). This has been taken as evidence for a particular difficulty with attention shifting in these disorders compared to, for example, individuals with moderate learning difficulties who show similar patterns of errors as ASD participants up to the ED stage, but less pronounced difficulties there (Hughes et al, 1994). However non-clinical populations have also been found to struggle with the ED stage of the

IEDS task, suggesting that such difficulties are not specific to ASD or Schizophrenia (Pantelis et al, 1999). Moreover, there are considerable differences in performance among samples from the same clinical populations. For example, Pantelis et al (1999) reported only around 25% of adults with Schizophrenia successfully completed all nine stages of the IEDS task, while Jazbec et al (2007) found around a 70% successful completion rate for the same clinical group.

Recent work has cast some doubt on the account of attention shifting difficulties in individuals with ASD and Schizophrenia. For example, Yerys et at (2009) found no difference between ASD and comparison groups in the number of errors made at the ED stage, and an age stratified analysis by Chen et al (2016) found a difference in performance between ASD and comparison participants in 8-12 year olds, but not 13-18 year olds, and even for the younger groups there was no difference in errors made at the ED stage. Alternative perceptual attention shifting paradigms, such as Smith et al (2019), have also failed to find consistent differences in performance between ASD and comparison groups. Similarly, in relation to Schizophrenia, a recently proposed theory of attentional filtering, dubbed 'hyperfocussing' (Luck et al 2019a; 2019b), suggests that differences in attention compared to comparison groups may be more subtle than previously believed, with a narrower but more intense focus that can sometimes provide improved performance, particularly in visual attention (Luck et al 2019b). While we might still expect that a 'narrow' attentional focus would impair extradimensional shifting, it is unclear whether hyperfocussing would predict impaired performance in the IEDS task for people with Schizophrenia. One possibility is that hyperfocusing might improve performance up to the ED stage, but not at the ED stage. Overall, these findings suggest either that our assumption of an attention shifting deficit in ASD and Schizophrenia requires further examination, or that our

understanding of the drivers of performance on the different stages of the IEDS task is incomplete.

The IEDS task is complex, and the data it generates are similarly complex. A key challenge when interpreting the findings of the IEDS task is that we lack an account of the process by which participants arrive at a decision on each trial. This makes it hard to answer questions about the reason for the relative performance of patients with Schizophrenia, ASD, or other conditions. The aim of this paper is to introduce a simple computational model of performance in the IEDS, and to demonstrate that this model can help us draw conclusions about the reasons for diminished performance in ASD, Schizophrenia and other conditions. The motivation for introducing such a model is to allow the exploration of competing effects of switching attention between and within dimensions and overall learning in determining performance.

Constructing formal models is particularly useful in cases such as these, since their primary aim is to *explain* performance, rather than simply *describing* it (van Rooij & Baggio, in press). Thus using a model allows us to bridge the gap between differences in performance between different clinical groups (which is merely a description of the problem) and differences in the underlying rates of learning and attention switching (which go some way towards providing an explanation). We will also show how the model can be used to compare performance of groups when the details of the task, or the way the results are reported, vary between experiments, helping to bring some clarity to a currently rather disparate literature.

The rest of this paper is structured as follows; We begin with an outline of the IEDS task, followed by a brief discussion of some of the controversy around its interpretation. Next we give an outline of the model and discuss, with the aid of some simulations, the extent to which the

effects of attention switching and general learning rate can be distinguished in the data. We then proceed to fit the model to a previously unreported data set of individuals with ASD and matched controls, and finally to some previously reported data from a wider set of clinical groups. We finish with a discussion of the findings. Extra modelling details can be found in the Supplementary Materials.

The IEDS Task

Let us give a brief overview of the IEDS task; a full description can be found in Downes et al., (1989). The basic set up is a discrimination learning task, where participants have to select which of two presented stimuli is 'correct'. Most of the time, the stimuli vary on two dimensions, purple shapes and white lines, and only one of these dimensions is relevant for identifying the correct stimulus. Examples of the sorts of stimuli used for the various stages of the task are shown in Figure 1.

For the first stage the stimuli consist of a single dimension (SD), either purple shapes or white lines, and participants have to learn which of the two options is 'correct'. Feedback is provided on every trial. If a participant makes six correct responses in a row[1], they progress on to the next stage. If a participant does not progress after 50 trials of a given stage then the experiment terminates.

At the second stage the 'correct' option is now reversed (SDr), so the previously incorrect line or shape is now the correct option. At the third compound discrimination (C-D) and fourth (CD) stages the second, dimension is introduced, first spatially separated from the original dimension, and then overlapping. At this point this second dimension (white lines in the example

in Figure 1) is irrelevant, and the correct response is still determined by the first dimension (purple shapes in Figure 1). At the fifth stage (CDr) the correct option reverses again.

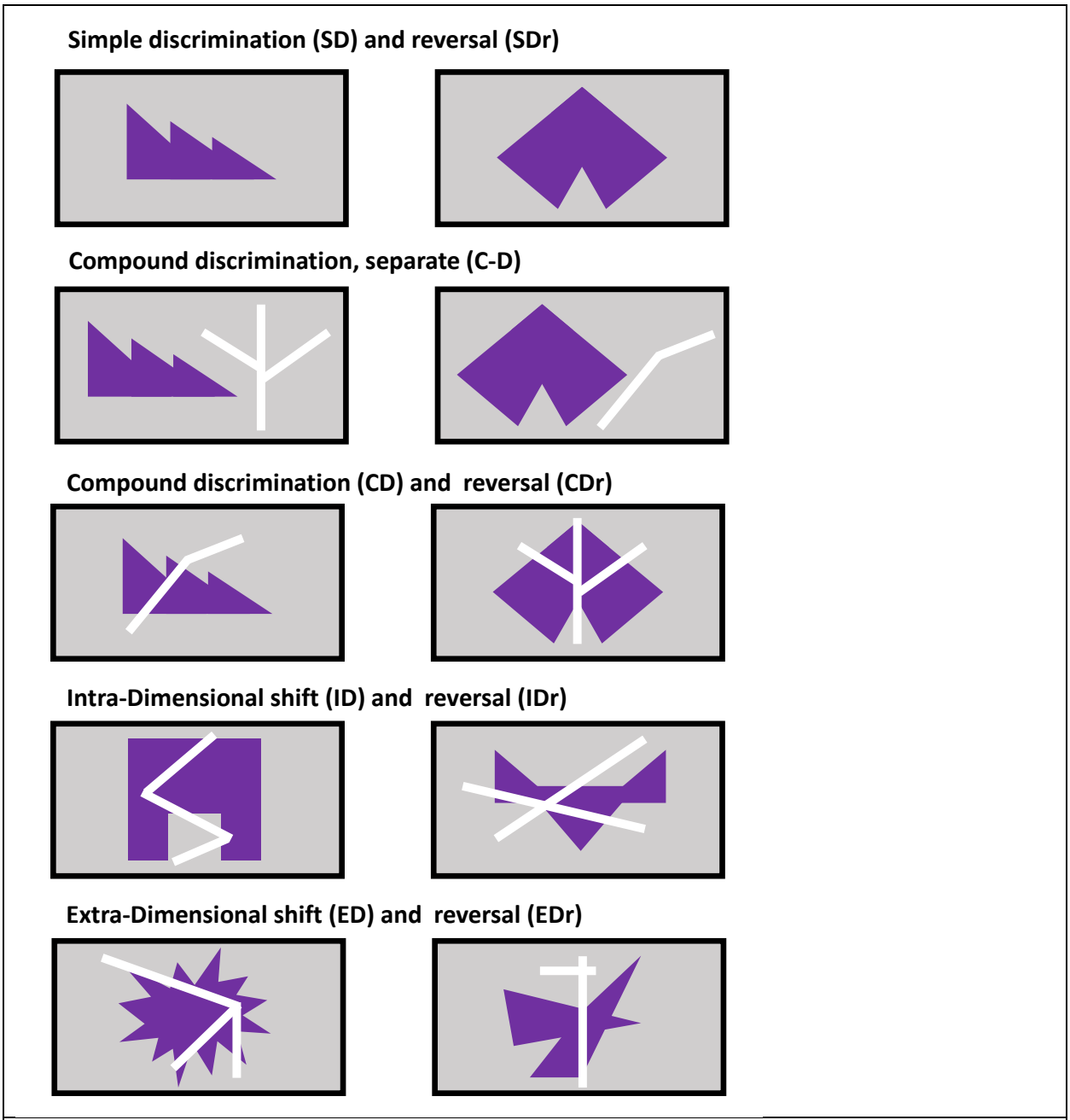


Figure 1: Examples of the sorts of shape and line stimuli used in the different stages of the IEDS. This figure is redrawn based on the examples given in Downes et al (1989).

The sixth (ID) stage is the Intra-dimensional shift. New exemplars of the stimuli dimensions (purple shapes and white lines) are introduced, but the relevant dimension remains

the same. The seventh stage (IDr) is another reversal. The eighth stage (ED) is the Extra-dimensional shift. New exemplars are again introduced, but now the relevant dimension switches to the previously irrelevant one, e.g. from the purple shapes to the white lines. The final ninth stage (EDr) is another reversal.

From the point of view of assessing attention switching ability, the key comparison is between performance in the sixth (ID) and eighth (ED) stages. In both of these stages new exemplars are introduced, so that the previously learnt information about which stimuli are correct is lost. However at the ID stage the higher order information, about which dimension is diagnostic, is still valid, but at the ED stage it ought to be impossible for participants to progress by attempting to learn the correct classification based on the old, now irrelevant, dimension. Success at the ED stage therefore requires switching attention between dimensions.

It is useful to note that the probability of successfully completing a stage by random guessing is quite high. Indeed, assuming a criterion of six correct responses in a row the success rate under guessing is around[2] 31%. This is important to keep in mind when assessing performance.

Problems with Interpretation

A difficulty when interpreting the results of the IEDS task comes from the fact that there is no consistent standard for reporting results. Reporting the attrition rate (i.e. percentage of participants successfully completing each stage) is relatively common but not universal (e.g. Chen et al (2016) only reported average number of stages completed). Reporting trials or errors to criterion for some or all stages is also common, although some studies only report values for the ID and ED stages, and even when all stages are included they are sometimes grouped into

four types, ‘Discrimination’ (SD, C_D, CD), ‘Reversal’ (SDr, CDr, IDr, EDr), ‘Intradimensional’ (ID) and ‘Extradimensional’ (ED). There are also inconsistent standards for how to deal with participants who do not make it to a given stage, with two strategies, either to ignore these participants, or to assume they would have taken 50 trials and made 25 errors on any stage they did not actually see. Finally, although the version of the task we described above required six correct guesses in succession to progress to the next stage, Hughes et al (1994) used a modified version requiring eight correct guesses to progress.

Taken together, these features make it hard to compare published results on the IEDS task in clinical populations. For example, the IEDS task has been taken by groups with ASD, and Schizophrenia, and both groups show reduced performance compared to comparison groups. It would be reasonable to ask whether one group has more difficulty with attention switching, however, the studies we have been able to find on these two populations do not report the same kinds of performance data, making a direct comparison difficult. To complicate matters further, as we shall see later on, it can be hard to separate the effects of reduced attention switching from a general reduction in learning rate.

The natural solution to this problem of interpretation is, we argue, to construct a model of the process of learning that happens in this task. The use of formal models is particularly important where there are multiple interacting factors which can affect performance, since intuition can often be an unreliable guide (Guest & Martin, 2020). Such a model would let us pick apart the relative impact of general learning impairments from a specific attention switching difficulty (i.e Jazbec et al., 2007). It would also provide a way to assess performance in a way that is independent of the precise details of the task or the format of the reported data.

Outline of the Model

We constructed a model of performance on the IEDS task following the same principles as a recent sequential learning model of the WCST by Bishara et al (2010). Both the WCST model and our model follow in a tradition known as cognitive psychometrics (Batchelder, 1998), that is, the purpose of the models is to allow us to categorise participants by estimating best fitting parameters, and to that end it is important that these models be simple and easy to fit to data, rather than accurate representations of real cognitive processes. This class of models has seen successful application in the Iowa Gambling Task (Busemeyer & Stout, 2002), the Balloon Analog Risk Task (Wallsten, Pleskac, & Lejuez, 2005), and the Go- No-Go Task (Yechiam et al., 2006).

The primary aims of the model are firstly to allow the differing effects of learning rate (both from positive and negative feedback) and attention switching to be compared and distinguished, so that we can attribute differential performance between controls and clinical groups to changes in one or the other of these parameters, and secondly to allow us to compare performance between groups when the reported data, or even the task details, are not identical. Note that we are not interested in comparing different possible models here, since we know of no other model which could be applied to this task. Instead we are interested in using our model as a tool to help us understand the drivers of performance in different clinical groups. If alternative models, which account for performance in different ways, are eventually produced, then we could address the issue of which model is ‘better’, but for the present purposes it is enough that the model captures the important features of the data.

Full details of the model are given in the Supplementary Materials, and code to simulate and fit the model to data is available online. Here we give an accessible overview of how the model works so that readers less interested in the maths can appreciate the results.

The idea of the model is that attention is placed on each of the four features (two shapes and two lines) that make up the two presented stimuli. One stimulus is then chosen based on the sum of the attention weights of the constituent features. When feedback (either positive or negative) on the choice is received, attention is shifted from the features which made up the incorrect choice, to the features that made up the correct one. Some attention is shifted between features of the same dimension (e.g. purple shapes), but some also shifts between dimensions, from the purple shape that made up the incorrect stimuli to the white line that made up the correct one, etc. The degree to which attention shifts between dimensions is determined by the attention shifting parameter f, and if $f = 0$ attention only shifts within a dimension. The overall rates at which attention is updated following positive or negative feedback are determined by the learning from reward (positive feedback) and learning from punishment (negative feedback) parameters r, p.

To successfully complete the task requires both good learning from the feedback obtained, and also an ability to switch attention between dimensions at the ED stage. In addition, there is a requirement that participants can sustain focus on the task. We assume for the purposes of model fitting that this ability to sustain focus does not vary between groups, however, particularly given the known co-morbidity between ASD and Attention Deficit Hyperactivity Disorder (e.g. Leitner, 2014), it is worth exploring this assumption further. We do this in the Supplementary Materials, and we show the results of varying the focus in Figure S2. In sum, a reduction in ability to sustain focus on the task does lead to reduced performance in the IEDS

task, but the pattern of data seems qualitatively different from that observed in any of the clinical groups studied here. It would be interesting to explore this further in future work.

Crucially according to the model there are therefore three independent ways in which a participant can struggle with the task; they can fail to consolidate attention on the correct feature following positive feedback, they can fail to switch attention between features following negative feedback, or they can fail to switch attention between dimensions at the ED stage. Actual participants may display some combination of these.

Simulations

We want to explore the behaviours the model displays for various parameter choices. Partly this is to demonstrate that the model does capture typical patterns seen in clinical data, for example selective difficulty at the ED stage. However, we are also interested in using the model to explore the claim that the IEDS task is able to pick apart the varying effects of attention switching vs overall learning rate. This will form an important starting point for fitting the model to real data.

We begin with a comparison between a 'control' set of parameters and a set with reduced attention switching. The 'control' set has good rates of learning from reward and punishment $r = p = 0.5$, and a moderate attention switching parameter, $f = 0.4$. The other set has the same rate of learning from reward and punishment, but a reduced attention switching parameter, $f = 0.1$. Naively, we should expect that the reduced attention switching parameter will result in performance comparable to the 'control' up to the ED stage, since it is only this stage that requires attention switching. Results are shown in Figure 2, where we compare the two parameter sets for 1,000 simulations in some of the typical ways seen in the literature. In Figure

S1 in the Supplementary Materials we present a visualisation of the way in which attention switches between features during the task. Overall, for these parameter values the attention switching parameter does selectively influence performance in the ED stage, which seems like a good endorsement of the model, and indeed of the IEDS task.

However, our intuition that the attention switching parameter selectively influences performance at the ED stage is problematic. The problem is that the model, or participant, doesn't know that no attention switching is required until the ED stage. Faced with negative feedback at, say, the CD stage, a participant may respond by switching attention to the, as yet, irrelevant dimension. It therefore seems plausible that high values of the attention switching parameter might result in worse performance up to the ED stage. We demonstrate this in Figure 3 with two parameter sets, both with lower rates of learning from reward and punishment ($r = p = 0.15$), and one with a low value of $f = 0.1$ and one with the maximum value of $f = 1.0$.

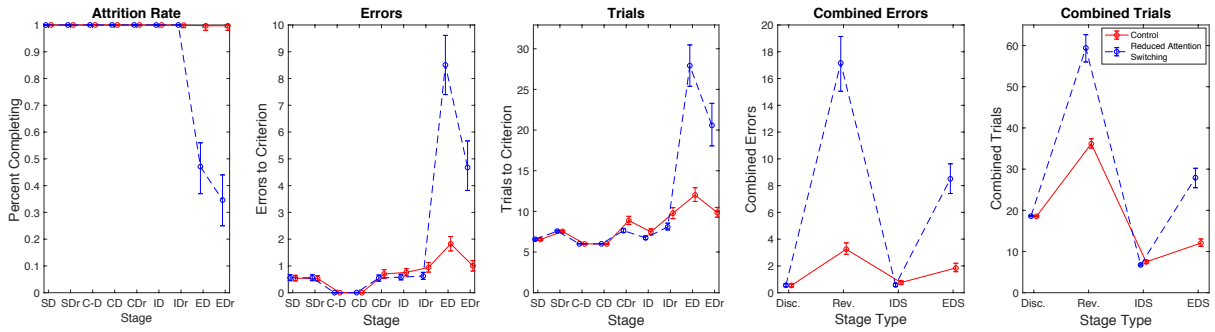
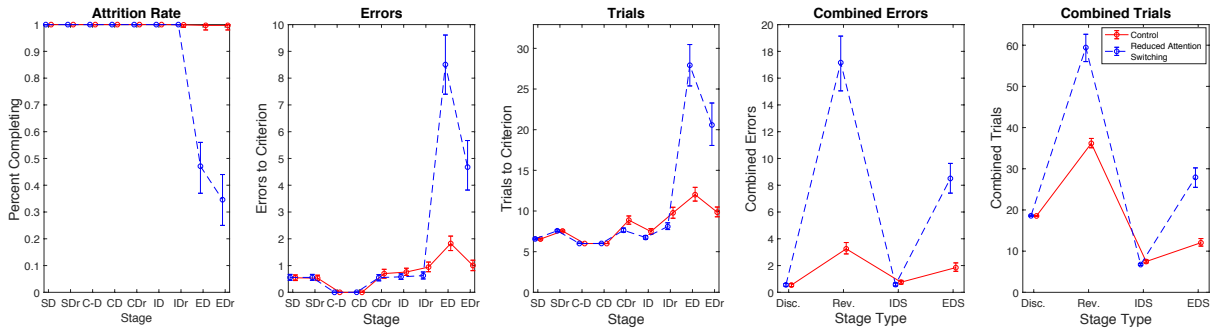


Figure 2. The effect of changing the attention switching parameter, *f*, on performance for fixed $r = p = 0.5$. Comparison of statistics for 1000 model simulations with the 'Control' (red, solid) and reduced attention switching (blue, dashed) parameters. Error bars are bootstrapped 95% confidence intervals assuming a group size of 100.



From left to right the panels show Attrition Rate, ie percentage of participants successfully completing each stage, Errors and Trials to criterion for each stage, and Errors and Trials to criterion summed for the four types of stages, discrimination, reversal, interdimensional shift and extradimensional shift. For the first three panels results are calculated only for those simulated participants actually reaching that stage, whereas for the final two panels we have adopted the standard convention of assigning 50 trials and 25 errors to any stage a participant failed to reach. The parameter sets produce similar behaviour until the ED stage, at which point the lower value of f hurts attention switching, with an increased number of errors and trials to criterion, and many simulated participants failing to complete the experiment.

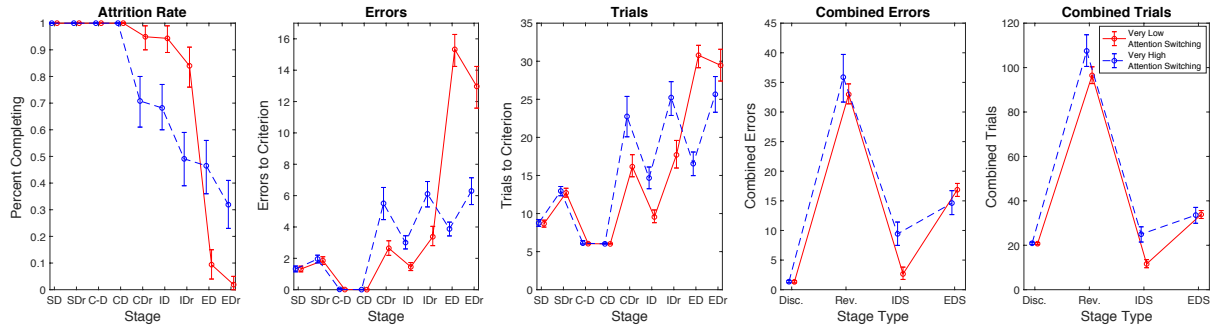


Figure 3. Exploring the expected performance for $r = p = 0.15$ and $f = .1$ (red, solid line) and $f = 1.0$ (blue, dashed line). Error bars are bootstrapped 95% confidence intervals assuming a group size of 100. Although a value of f = 1 clearly benefits the model at the ED stage, it diminishes performance earlier in the experiment, in particular at the reversal stages CDr and IDr. Panel descriptions are as for Figure 2.

As expected, although a high value of f does benefit the model at the ED stage, it diminishes performance up to that point. In particular it seems to reduce performance at reversal

stages, since participants always make errors at these stages, and the high rate of attention switching causes the participant to incorrectly make an extra-dimensional shift.

Finally, and most importantly for the rest of this paper, although poor performance at the ED stage can be the result of poor attention switching, it can also be caused by poor learning from punishment. This is again somewhat intuitive, since attention switching is a form of learning, so overall impairments in learning can impede successful attention switching too. In Figure 4 we compare performance between impaired attention and impaired learning from punishment. The two parameter sets produce generally different behaviours, but they agree almost exactly at the ED stage, demonstrating that poor learning from punishment can imitate impaired attention shifting, at least when only looking at the ED stage.

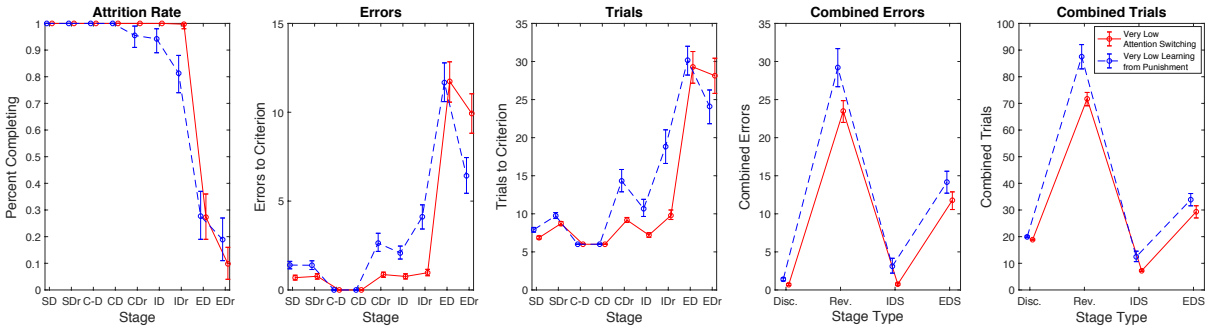


Figure 4. Comparing performance for r = 0.3, p = 0.3, f =0.08 (red, solid line) and r=0.3, p=0.08, f=0.3 (blue, dashed line), (note the symmetry in parameter values). Error bars are bootstrapped 95% confidence intervals assuming a group size of 100. Although the parameters produce generally different behaviours, they agree almost exactly at the ED stage. Panel descriptions are as for Figure 2.

Overall then, the simulations paint a mixed picture; for certain learning rate values, attention switching does essentially selectively influence performance at the ED stage, but for other learning rate values, the influence is not selective, and there is a trade-off between performance at the ED stage and performance at earlier stages. In addition, poor performance at

the ED stage is not by itself diagnostic of impaired attention switching but can also result from poor general learning.

We can also say something about the sort of performance we expect to be characteristic of poor attention switching ability, as opposed to general poor rates of learning. Specifically, excellent performance up to the ED stage, followed by difficulties, is likely indicative of poor attention switching. However, a more general difficulty throughout the IEDS task, culminating in severe problems at the ED stage, is more likely diagnostic of poor general learning, rather than an attention switching deficit. We shall see this in more detail when we come to fit the model to real data, but already we can see the model producing insight that would be difficult to have gleaned by other means.

Fitting the model to real data

The next challenge for the model is to see how well it can capture real data. There are two possible ways to approach this, depending on the data we have available. In this section we will work with a complete data set for a large number of individuals, where we have full information about the stimuli presented and decisions made for each trial. In the next section we will show how to extend model fits to situations where we only have summary data, as typically reported in papers using the IEDS.

The data used in this section is based on previously unpublished data from 104 individuals recruited at City, University of London. Fifty-two adults with ASD (42 men, M_{age} = 43.71 years, range: 25-65 years) were closely matched to 52 typically developing (TD) participants (40 men, M_{age} = 42.52 years, range: 21-65 years) on chronological age, gender, Verbal Intelligence Quotient (VIQ), and Performance IQ (PIQ) as measured by the third or

fourth edition of the Wechsler Adult Intelligence Scale (WAIS-III^{UK} or WAIS-IV^{UK}; The Psychological Corporation, 2000; 2008; see Table 1). Participants belong to a database of individuals who regularly take part in research at the Autism Research Group at City, University of London. Prior to testing all ASD participants provided confirmation that they had received their clinical diagnosis through the UK's National Health Service according to DSM-IV-TR criteria (American Psychiatric Association, 2000). TD participants did not report a personal or family history of a psychological or neurodevelopmental disorder and did not take any psychotropic medication. All participants filled in the Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) as a means of sample description in terms of autism-like traits. All TD individuals had scores below the suggested cut-off core of 26 on the AQ (Woodbury-Smith, Robinson, Wheelwright & Baron-Cohen, 2005). ASD individuals obtained significantly higher scores compared to TD individuals on this instrument (see Table 1).

Participants were reimbursed for their time according to standard university rates and their travel costs were paid. Data collection was approved by the Psychology Department of City, University of London's ethics committee and the procedures used in this study followed the ethical guidelines set out by the British Psychological Society and were in accordance with the provisions of the Declaration of Helsinki.

	ASD (42m, 10f)		**TD (40m, 12f)**					
Measure	***M***	***SD***	***M***	***SD***	***t*(df)**	***p***	**Cohen's *d***	**CI**
Age (years)	43.71	12.45	42.52	12.25	0.49 (102)	.64	0.10	-0.29, 0.48
VIQ/VCI[a]	113	16.7	113	15.6	0.31 (102)	.76	0.06	-0.32, 0.44

PIQ/PRI[b]	107	15.9	108	15.0	0.27 (102)	.89	0.05	-0.33, 0.44
AQ[c]	34.92	7.58	14.10	5.55	15.98 (93.50[d])	<.001	3.13	2.53, 3.68

Table 1. *Descriptive statistics for individuals with Autism Spectrum Disorder (ASD) and Typically Developing (TD) controls. Note.* [a]VIQ - Verbal IQ (WAIS-III^{UK}) or VCI - Verbal Comprehension Index (WAIS-IV^{UK}). [b]PIQ - Performance IQ (WAIS-III^{UK}) or PRI - Perceptual Reasoning Index (WAIS-IV^{UK}). [c]Autism-Spectrum Quotient. [d]Equal Variances not assumed.

Details of the model fitting are given in Supplementary Materials. The fitting process produces posterior estimates for the parameters, but we will focus our analysis on just the means for each parameter and participant. Our main question is whether the best fitting parameter estimates predict behaviour and, ultimately, AQ score and diagnosis.

Results

In Figure 5, we plot descriptive data for the TD vs ASD group, in the ways commonly seen in the literature. What is striking is that there does not appear to be a separation between the groups at any stage of the task. Indeed, regressing the group against the difference in performance at the ED and ID stages, does not produce a significant regression ($p > .05$)[3]. This is also true when regressing AQ score against the difference in performance at the ED and ID stages. Unsurprisingly then, attempting to predict to which group a participant belongs, based on their best fitting model parameters, r, p, f, does not produce a significant result ($p > .05$).

The best fit learning from reward and learning from punishment parameters are strongly correlated $\rho_S(104) = .71,\ p{\sim}10^{-18}$. However learning from reward and learning from punishment also both correlate to a lesser degree with the attention switching parameter,

$\rho_S(104) = .50,\ p \sim 10^{-8}$, and $\rho_S(104) = .26,\ p = .007$ respectively. These results are essentially unchanged when looking at the TD and ASD groups separately, although the correlations are all stronger for TD vs ASD participants. Note that increases in all three parameters will tend to produce 'better' performance in a way that may not be totally distinguishable at the level of individual participants, especially given limited amount of data available per participant.

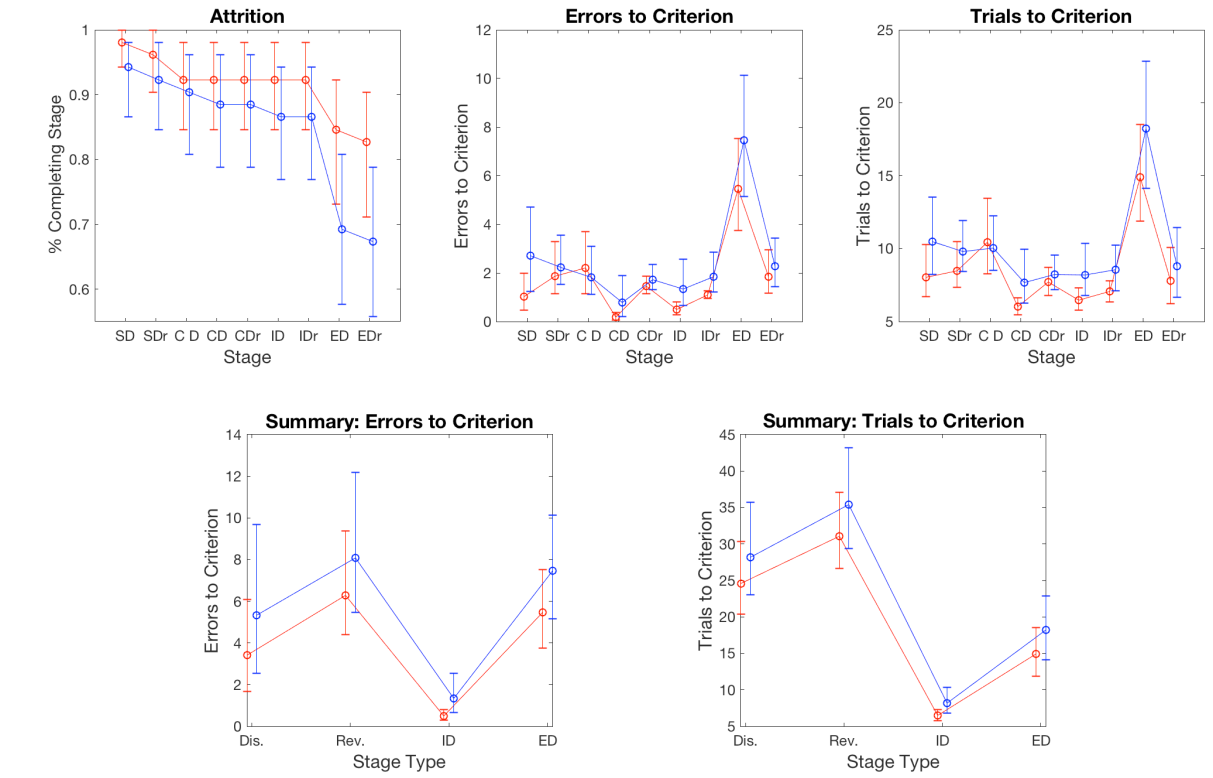


Figure 5. Summary statistics for the full data from City, University of London. In all graphs the red solid lines represent controls and the blue dashed lines participants with a diagnosis of ASD. Error bars are bootstrapped 95% confidence intervals. Top row shows attrition (left), errors to criterion (centre) and trials to criterion (right). The bottom row shows summed errors (left) and trials (right) for four categories of stage, discrimination, reversal, intradimensional shift and extradimensional shift.

However, the best fitting model parameters do predict errors at ED minus errors at ID, $F(2, 101) = 21.36,\ p < .0001,\ R^2 = .39$, trials at ED minus trials at ID, $F(2,101) = 19.35,\ p < .0001, R^2 = .37$, and number of stages completed, $F(2, 101) = 12.79,\ p < .0001, R^2 = .28$. This makes the conclusion that the model explains individual differences in performance well, but performance is not related to AQ score or diagnosis, hard to escape.

Interestingly, when looking at errors and trials to criterion at the ID and ED stages, the greatest single predictor is the attention switching parameter, which by itself explains 30.1% of the variance in the difference in errors at the ED and ID stages, and 29.8% of the variance in the difference in trials at the ED and ID stages. However, when looking at the number of stages completed, the attention switching parameter is not a significant predictor by itself, and the best single predictor is the learning from punishment rate, accounting for 23.8% of the variance. In other words, poor attention switching leads to selective difficulty at the ED stage relative to the ID, but progression through the task also depends strongly on learning from punishment. This echoes some of the findings of the simulations, where we saw that poor learning from punishment could cause difficulties even with unimpaired attention switching. We can also clearly see from the data that neither the ASD nor the control group has the signature of the pure attention switching deficit seen in the simulations.

We can further explore the link between learning and task performance by examining the relationship between the IQ of participants, task performance, and model parameters. Neither PIQ or VIQ predict relative difficulty at the ED vs ID stage in terms of errors or trials taken. However VIQ predicts the total number of stages completed, $F\ (2, 101) = 15.17,\ p = .0002,\ R^2 = .13$, as does PIQ, $F(2, 101) = 22.12,\ p < .0001,\ R^2 = .18$. (For PIQ this is true when looking at the ASD and comparison groups separately, for VIQ this is true for ASD but not

comparisons). This is nicely explained by the fact that VIQ and PIQ are significantly associated with the learning from punishment parameter, p, $F(2, 101) = 16.42,\ p = .0001,\ R^2 = .14$ and $F(2, 101) = 12.05,\ p = .0008,\ R^2 = .11$ respectively, but that neither are associated with the attention switching parameter. (These conclusions also hold when analysing the ASD and comparison groups separately.) This suggests the learning from punishment parameter is sensitive to general intelligence, and this is a key determinant of the number of stages completed.

Summary

Results from this large data set suggest little relationship between performance on the IEDS task and AQ score or a diagnosis of ASD. The total number of stages completed seems instead to depend on VIQ/PIQ. The modelling supports these conclusions, with the expected relationship between selective difficulty at the ED stage and the attention switching parameter, and a significant relationship between PIQ/VIQ and the learning from punishment parameter. The model also helps explain why number of stages completed is better predicted by VIQ/PIQ than AQ.

Individuals in this data set are well matched on IQ, and the ASD group have a diagnosis based on DSM-IV. If individuals in this sample with ASD do have attention switching deficits, this is not apparent either from the data or from the modelling.

Fitting the model to Summary Data

Having explored performance in the IEDS task for a particular data set of ASD vs matched comparison participants, we now want to broaden our study to previously collected data, and also data from other clinical groups. Our aim is first, to explore the extent to which our

key finding – no evidence for impaired attention switching in our ASD sample - generalises to other groups of individuals with ASD and to other conditions. Our second aim is to explore whether there are differences between ASD and other clinical groups in this task.

In the last section we discussed fitting our model to the full data from a group of participants. Crucially we had access to both the stimuli presented and the responses made for every trial, and we can use our model to compute the probability that a participant would make a particular response, given the stimuli shown on that trial, and the history of trials and responses up to that point.

However, this level of detail is never reported in the literature on the IEDS task. What is typically reported is average data for the comparison and clinical group(s) in a number of different forms; either attrition rates, errors/trials to criterion for all stages, or sometimes summed errors/trials to criterion for the four different types of stage. Despite only having summary level data from previously reported studies, our aim remains to examine how well our model captures behaviour.

Fitting the model to this summary data presents a different challenge from fitting when the full data is available. We shall make use of the techniques of approximate Bayesian computation to allow us to estimate best fitting parameters in a way which bypasses computation of an explicit likelihood (Turner & van Zandt, 2012). Full details are in the Supplementary Materials.

These techniques allow us to fit our model to reported data from four previously published studies using the IEDS task. These studies were selected to cover a reasonably broad range of dates, and types of data reported, but we make no claim that these are representative studies. Basic information about the different studies can be found in Table 2. Two of the studies

compared controls to individuals diagnosed with Schizophrenia (Jazbec et al, 2007; Pantelis et al, 1999), two studies compared controls to participants with Autism Spectrum Disorder (Ozonoff et al, 2004; Hughes et al, 1994). One study also included a group with moderate learning difficulties (Hughes et al, 1994), and another included a group with prefrontal cortex lesions (Pantelis et al, 1999).

Study.	Control Group	Clinical Group 1	Clinical Group 2	Data Reported[a]
Hughes et al (1994)	47 children, mean age 8	35 children with ASD, mean age 13	38 children with moderate learning difficulties, mean age 13	Attrition Rate. Errors to criterion at the ID and ED stages.
Pantelis et al (1999)	18 adults, mean age 40	24 adults with Schizophrenia, mean age 48	22 adults with prefrontal cortex lesions, mean age 45	Attrition Rate, Errors and Trials[b] to criterion, conditional on passing.
Ozonoff et al (2004)	70 children, mean age 16	79 children with ASD, mean age 16	-	Trials and Errors to criterion at ID, IDr ED & EDr stages.
Jazbec et al (2007)	26 adults, mean age 41	36 adults with Schizophrenia, mean age 35	-	Attrition rate. Trials to criterion, conditional on passing.

Table 2. Summary of the four studies fit in this section.

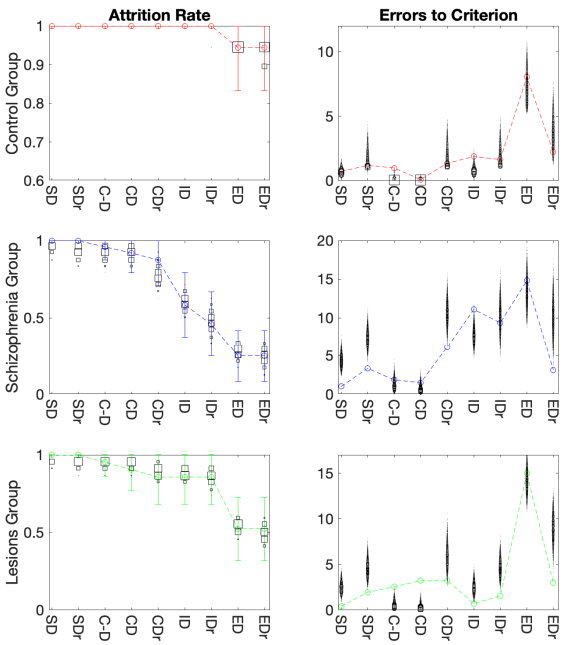
[a] Pantelis et al (1999) and Jazbec et al (2007) also report unconditional data, but this doesn't add any extra information.
[b] Fits were done to attrition and error data only.

Although there is some overlap in the clinical groups, it is difficult to directly compare the performance of the different groups between experimental studies. For example, although Hughes et at (1994) and Ozonoff et al (2004) both included a group of children with ASD, there

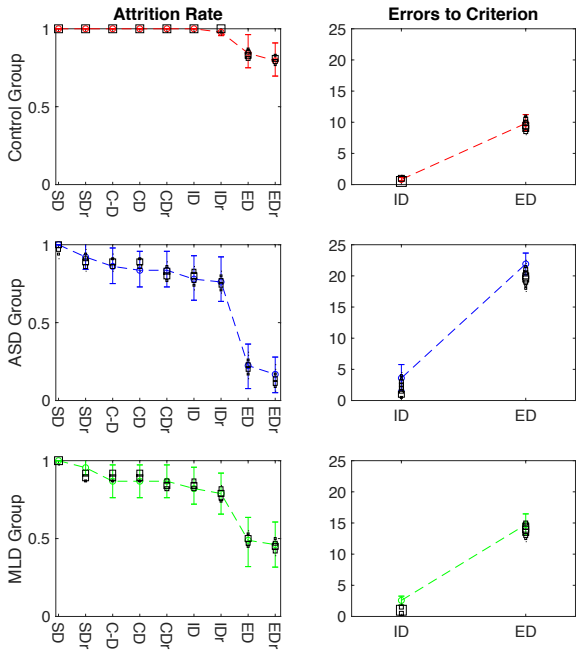
is almost no overlap in the type of data reported. Part of the advantage of a model-based approach is that it can produce parameter estimates for the different groups that can be easily compared across experiments, even though the data reported may be very different.

Figure 6 shows the model fits to the data reported in each of the studies in Table 2. In each case fits are generally good, with only small mismatches in a few places. In particular, performance at the ID and ED stages is well captured in all fits, giving us confidence that the model captures the different groups' performance well.

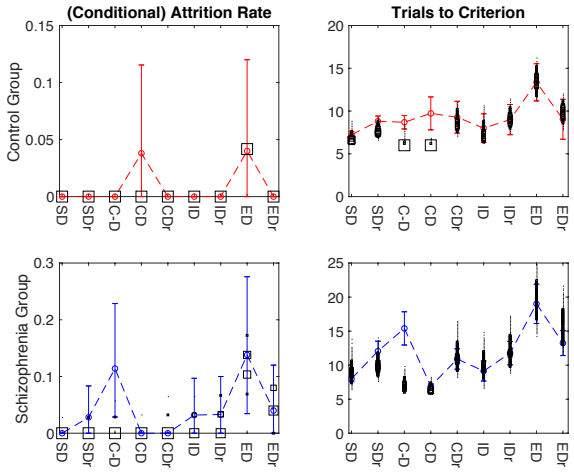
The means and 95% Highest Density Intervals (HDIs) for the model parameters for each group are shown in Table 3. The HDIs for the learning parameters in the control groups are very wide due to a ceiling effect – even moderate learning rates are enough to produce good accuracy in all stages as long as the attention switching parameter is not too small, and so the model is relatively insensitive to the exact values in these groups. Posteriors for the attention switching parameter, and for the other groups, are generally much tighter, reflecting the sensitivity of the model to the exact degree of difficulty in either learning or attention switching.



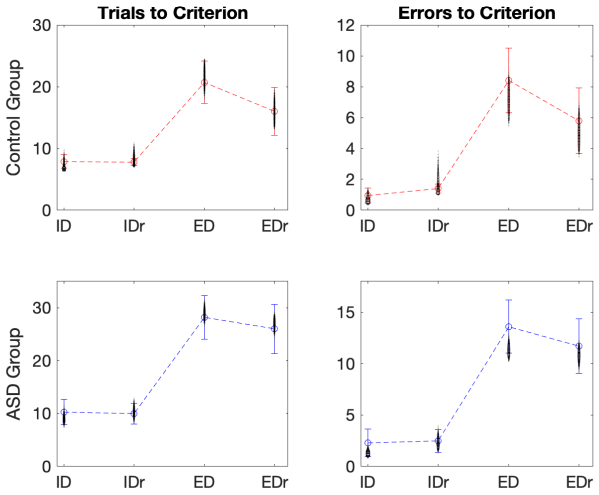
(a) Data and fits for Pantelis et al (1999). Error bars for attrition rates are bootstrapped 95% confidence intervals, it was impossible to estimate the variance for errors to criterion from the original paper.



(b) Data and fits for Hughes et al (1994). Error bars for attrition rates are bootstrapped 95% confidence intervals, error bars for errors to criterion are estimated from the original figures.



(c) Data and fits for Jazbec et al (2007). Error bars for conditional attrition rates are bootstrapped 95% confidence intervals, error bars for trials to criterion are ±1SE.



(d) Data and fits for Ozonoff et al (2004). Error bars for trials and errors to criterion are ±2SE.

Figure 6. Data and fits for each of the four experiments analyzed in this section. Dashed lines are data and black boxes are histograms of the posterior probability distribution produced by the fitting process.

Given the good overall match between model and data, we can compare the estimates for the three parameters between fits, to answer our central question about whether and how the various clinical groups differ from each other, and from the controls. We do this by computing posterior difference distributions, as described in the Supplementary Materials and presented in Figures S3-S5. Here, we will present the most theoretically important comparisons, obtained by examining a combination of Bayes factors (Kass & Raftery, 1995) for the hypotheses of difference vs no difference, and 95% HDIs of the posterior difference distribution. Bayes Factors indicate the relative evidence for the hypotheses, with values of >3 as weak evidence and >10 as strong evidence for one hypothesis over another (Kass & Raftery, 1995). The 95% HDIs give the region where we are 95% confident the true difference lies (the interpretation of HDIs is simpler than for regular confidence intervals.)

First, we check consistency between the control groups across experiments. For the learning from reward and learning from punishment parameters we have no evidence of differences between the control groups. For the attention switching parameter the data from Ozonoff et al (2004), Hughes et al (1994) and Pantelis et al (1999) are consistent, but the data from Jazbec et al (2007) seems to have a slightly different pattern, producing a higher estimate for the attention switching parameter (.28 vs .13-.14). This is consistent with the control group in Jazbec et al (2007) seeming more likely to successfully complete the ED stage, taking fewer trials to do so than the other control groups. In sum, data from the control groups can be compared with the help of the model fitting, even though the data is not directly comparable, and the groups seem fairly consistent.

Study	**Pantelis et al (1999)**			**Hughes et al (1994)**			**Jazbec et al (2007)**		**Ozonoff et al (2004)**	
Group	Control	SZ	Lesions	Control	ASD	MLD	Control	SZ	Control	ASD
r	.543	.002	.023	.839	.105	.885	.487	.025	.446	.053
HDI	[.021, .985]	[.000, .028]	[.000, .120]	[.414, .998]	[.000, .974]	[.574, .995]	[.131, .944]	[.000, .207]	[.028, .972]	[.000, .186]
p	.622	.155	.254	.335	.154	.028	.521	.485	.670	.508
HDI	[.265, .988]	[.114, .204]	[.162, .354]	[.093, .962]	[.016, .366]	[.017, .049]	[.325, .928]	[.269, .638]	[.308, .988]	[.327, .693]
f	.131	.307	.119	.143	.028	.142	.283	.215	.130	.106
HDI	[.069, .234]	[.127, .594]	[.053, .212]	[.085, .234]	[.002, .068]	[.096, .200]	[.212, .380]	[.146, .314]	[.084, .196]	[.073, .149]

Table 3: Means and 95% HDIs of the parameter posteriors for each of the groups fit here. For details about significant comparisons see the main text and Supplementary Materials. Recall r is the learning from reward parameter, p is the learning from punishment parameter, and f is the attention switching parameter.

	Between Groups	**Comparison with control groups**
Learning from Reward	Strong evidence of equality	Mixed/Strong evidence < controls
Learning from Punishment	Strong evidence J > P	J: No evidence of difference P: Mixed/Strong evidence < controls
Attention Switching	No evidence of difference	No evidence of difference

Table 4: Summary of parameter comparisons for the Schizophrenia groups. J = Schizophrenia group from Jazbec et al (2007), P = Schizophrenia group from Pantelis et al (1999).

Strong evidence means all Bayes Factor >10, Weak evidence means all Bayes Factor >3. Mixed/Strong evidence means at least some comparisons have Bayes Factor >10, Mixed/Weak evidence means at least some comparisons have Bayes Factor >3. No evidence means no Bayes Factors exceed 3.

Turning to the two Schizophrenia groups' results, summarised in Table 4,we see that rates of learning from reward and punishment are lower for the Schizophrenia groups than the comparison groups, however we fail to see any evidence of a difference in attention switching rates between Schizophrenia and comparison groups. In other words, the modelling results do not support the idea of a specific attention switching deficit for these two Schizophrenia groups.

It is also important to note that there are significant differences between the two groups, both in performance and parameter estimates. In particular the estimated value of the learning from punishment parameter is significantly higher for the SZ group from Jazbec et al (2007) compared with Pantelis et al (1999), perhaps a reflection of the fact that the former had rather

higher mean WAIS-R score (91.3 vs 80.8), if we assume the relationship between IQ and learning from punishment we found in the City data generalises to SZ groups.

	Between Groups	**Comparison with control groups**
Learning from Reward	Weak evidence of equality	O: Mixed/Strong evidence < controls H: No evidence of difference
Learning from Punishment	Weak evidence O > H	O: No evidence of difference H: Mixed/Weak evidence < controls
Attention Switching	Weak evidence O > H	O: Weak evidence for equality[a] H: Mixed/Strong evidence < controls

Table 5: Summary of parameter comparisons for the ASD groups. O = ASD group from Ozonoff et al (2004), H = ASD group from Hughes et al (1994).

Strong evidence means all Bayes Factor >10, Weak evidence means all Bayes Factor >3. Mixed/Strong evidence means at least some comparisons have Bayes Factor >10, Mixed/Weak evidence means at least some comparisons have Bayes Factor >3. No evidence means no Bayes Factors exceed 3.

Notes: [a] Except for the control from Jazbec et al (2007).

Looking at the two ASD groups' results, summarised in Table 5, we have a more mixed picture. There is some evidence rates of learning from reward and punishment are lower for ASD compared with controls. However, while Ozonoff et al (2004) gives weak evidence that the attention switching parameter is the same as the controls, for Hughes et al (1994) we have some evidence of a lower attention switching parameter. In other words, the ASD group from Ozonoff et al (2004) looks much like the Schizophrenia groups, with no evidence for a specific attention switching deficit. However, the ASD group from Hughes et al (1994) do seem to have an

attention switching deficit, although this may be accompanied by a reduction in learning from negative feedback.

Similarly to the SZ groups, there is some evidence of differences in best fitting parameters between the ASD groups, although it is harder to know whether these reflect differences in the characteristics of the participants, as Ozonoff et al (2004) and Hughes et al (1999) do not give directly comparable assessments of IQ.

Finally, we turn to the other two groups, those from Hughes et al (1994) with moderate learning difficulties, and those from Pantelis et al (1999) with frontal lobe lesions. Here we can compare the groups against controls, and also against the ASD and SZ groups.

	Comparison with control groups	**Comparison with ASD/SZ groups**
Learning from Reward	No evidence of difference	Strong evidence > SZ Mixed/Strong evidence > ASD
Learning from Punishment	Strong evidence < controls	Strong evidence < SZ Strong evidence > Ozonoff et al (2004) ASD group Strong evidence of equality with Hughes et al (1994) ASD group[a]
Attention Switching	Weak evidence for equality[b]	No evidence of difference vs SZ Mixed/Strong evidence > ASD

Table 6: Summary of parameter comparisons for the Moderate Learning Difficulties group.

Strong evidence means all Bayes Factor >10, Weak evidence means all Bayes Factor >3. Mixed/Strong evidence means at least some comparisons have Bayes Factor >10, Mixed/Weak evidence means at least some comparisons have Bayes Factor >3. No evidence means no Bayes Factors exceed 3.

Notes: [a] However the posterior difference distribution is very irregular in shape, making the approximation leading to the Bayes Factors potentially unreliable.

[b] Except for the control from Jazbec et al (2007).

Results from Hughes et al's (1994) moderate learning difficulties group are summarised in Table 6. Overall, this group seems to have lower rates of learning from punishment compared with controls, but no specific attention switching deficit. Although the lack of a specific attention switching deficit for individuals with learning difficulties is similar to the Schizophrenia groups and one of the ASD groups, the model identifies different factors leading to impaired performance relative to controls, and there is strong evidence that the parameters are not the

same as for the ASD and SZ groups. It is impressive that the model is able to see this difference, especially as it is in line with predictions based on the relationship between IQ and learning from punishment seen in the previous section.

	Comparison with control groups	**Comparison with ASD/SZ groups**
Learning from Reward	No clear pattern	Mixed/Strong evidence for equality
Learning from Punishment	Mixed/Weak evidence < controls	Mixed/Weak evidence < ASD No evidence of difference vs SZ
Attention Switching	Mixed/Weak evidence for equality[a]	No evidence of difference vs SZ Mixed/Weak evidence for equality vs ASD

Table 7: Summary of parameter comparisons for the Frontal Lobe Lesions group.

Strong evidence means all Bayes Factor >10, Weak evidence means all Bayes Factor >3. Mixed/Strong evidence means at least some comparisons have Bayes Factor >10, Mixed/Weak evidence means at least some comparisons have Bayes Factor >3. No evidence means no Bayes Factors exceed 3.

Notes: [a] Except for the control from Jazbec et al (2007).

Finally, we examine the group with frontal lobe lesions from Pantelis et al (1999), results are summarised in Table 7. Overall, this group looks similar to the ASD and SZ groups, which is rather different from the group with learning difficulties.

Summary

Results from the modelling suggest that differences in performance in the IEDS task amongst individuals with ASD, Schizophrenia, Moderate Learning Difficulties or Frontal Lobe lesions are very unlikely to result from a particular difficulty with attention switching relative to matched controls. This may seem rather surprising given the fact that some of these groups show significant difficulties at the ED stage of this task. But this is precisely the value of such a computational model - we can pick apart the different influences of learning and attention switching in a way that is hard to do by looking at performance data by itself.

Equally, we can see that there are also differences between the two ASD and Schizophrenia groups in terms of best fitting parameter values. For the Schizophrenia group we can interpret this in terms of the difference in WAIS-R scores. For the ASD groups it is less clear how to understand this.

We can also say something about the different reasons for poor performance between the ASD and Schizophrenia groups. The Hughes et al's (1994) ASD group differ most markedly from the Schizophrenia groups in having a lower attention switching parameter. On the other hand, Ozonoff et al's (2004) ASD group differ from one Schizophrenia group in terms of having a higher learning from punishment parameter, and do not substantially differ in any way from the other group. This might be explained by the fact that Hughes et al's (1994) group was somewhat younger (mean age 13 rather than 16 years) than the group from Ozonoff et al (2004), and may be related to the work of Chen et al (2016) showing a difference between ASD and controls in younger but not older children, if we assume learning from punishment correlates with mental age in ASD (which is reasonable since it correlates with PIQ/VIQ).

It is also interesting that we can compare groups with Moderate Learning Difficulties and Frontal Lobe Lesions to the ASD and Schizophrenia groups. All four groups show impaired performance on the IEDS task relative to controls, but the model is able to conclude that the group with general learning impairments are distinct from the other three groups. Given the simplicity of the model it is impressive that it is able to make this distinction.

General Discussion

What can we conclude from this study? In a large and robustly matched sample of adults with ASD and controls we see considerable variation in IEDS task performance. Largely, this is well accounted for by the model we have developed, with the difference in errors made and trials taken at the ID and ED stages correlating with the specific attention shift parameter in the model, and a correlation between the learning from punishment parameter, the number of stages completed, and PIQ/VIQ. However, neither test performance nor the model parameters predict diagnosis or AQ score. In other words, variation in performance on the IEDS task seems to be completely unrelated to ASD.

We followed this up with an analysis of some previously published data sets from groups of participants with ASD, Schizophrenia, Moderate Learning Difficulties and Frontal lesions. While some of these groups do struggle in the IEDS task relative to controls, on the whole the modelling does not support the existence of a specific attention switching impairment for these groups. Instead, the model implicates reduced rates of learning from positive or negative feedback for most of these groups.

Although these conclusions are surprising, there are other reports of a failure to see specific attention switching difficulties in ASD groups. Yerys (2009) found a difference between

ASD and control groups in errors made at the EDr stage but not the ED stage, which would not be expected from our model given a reduced attention shifting parameter. The age stratified analysis by Chen et al (2016) found a difference in performance only in younger children, and again no difference in errors at the ED stage between ASD and controls. Instead Chen et al (2016) found a difference in the number of stages completed and errors made up to the ED stage for the younger group, and no differences for the older group. From our modelling we might expect this to result from a difference in learning from punishment which we saw was correlated with PIQ/VIQ. Interestingly, when looking at the data from Chen et al (2016) we see there is indeed a difference between the ASD group and controls in VIQ and full-scale IQ for the younger age groups, but not the older age groups. This finding points to the involvement of some sort of developmental process on which our modelling approach might help to shed some light. More systematic developmental studies could explore the relationship between IEDS performance, model parameters, and developmental level. A similar failure to see differences between IQ-matched Schizophrenia and comparison groups at the ED stage was noted by Leeson et al (2009).

Given the failure to observe a specific attention switching impairment in the ASD and Schizophrenia groups examined here, should we therefore conclude that attention switching difficulties are not associated with these conditions? Although this possibility merits examination, we think it is too bold. What seems more likely is that there are multiple notions of attention switching that might be linked to executive function, and the IEDS task measures a highly simplified one that is not clinically relevant. For Schizophrenia this may be related to the distinction drawn by Luck and Gold (2008) between ‘input selection’ and ‘rule selection’.

The IEDS task is a deliberately simple one, but some of these simplifications might affect the extent to which the task measures the useful notion of attention shifting. For example, unlike the WCST the IEDS task includes only two dimensions, so that attention shifting does not require searching for the new relevant dimension. Luck and Gold (2008, p. 34) argue that the attention deficits in Schizophrenia are related to "processes that guide attention to task-relevant inputs", and one might argue that a task with only two displayed stimuli and two feature types is too simple for any deficits to be apparent. Also, unlike more real-world decisions, success given the correct sorting rule is guaranteed, rather than probabilistic, and this consistent feedback may also make the ED shift easier and less realistic. It would be interesting to explore whether gradually increasing the complexity of the task in these ways leads to qualitatively different behaviours in clinical and comparison groups.

Finally, this analysis has, we hope, demonstrated the power of formal modelling in helping us understand performance in tasks like the IEDS. We have been able to generate insights that would have been impossible just from a statistical analysis of the performance data. The model was able to do this despite its relative simplicity, but future work could conceivably extend the scope of the model. One possibility would be to examine whether allowing the attention focus parameter to vary could capture the performance of other groups, such as those with ADHD. Another possibility would be to extend the model to cover similar tasks such as the WCST. We hope this will prove inspirational to others working with this class of tests.

Acknowledgements

Model fitting was performed on City, University of London's Solon High Performance Computing Cluster. We are grateful for this support. We are also grateful to Nathan Evans, for helpful discussions about parameter comparisons.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders-DSM-IV-TR (4th ed., text revision).* Washington, DC.

Barch, DM, Braver, TS, Carter, CS Poldrack, RA & Robbins, TW (2009). CNTRICS final task selection: executive control. Schizophrenia Bulletin, 35(1), 115-135.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high- functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31(1), 5-17.

Busemeyer, J.R., & Stout, J.C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*(3), 253–262.

Chen, S.-F., Chien, Y.-L., Wu, C,-T., Shang, C.-Y., Wu, Y,-Y. and Gau, S.S. (2016). Deficits in executive functions among youths with autism spectrum disorders: an age-stratified analysis. Psychological Medicine, 46, 1625-1638.

Downes, J.J., Roberts, A.C., Sahakian, B.J., Evenden, J.L., Morris, R.G., Robbins, T.W., (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: evidence for a specific attentional dysfunction. *Neuropsychologia* 27, 1329-1343.

Guest, O., & Martin, A.E. (2020). How computational modelling can force theory building in psychological science. https://doi.org/10.31234/osf.io/rybh9

Grant, D.A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in Weigl-type card-sorting problem. Journal of Experimental Psychology, 38, 404-411.

Hill, E.L. (2004) 'Evaluating the Theory of Executive Dysfunction in Autism', *Developmental Review* 24: 189–233.

Hughes, C., Russell, J. & Robbins, T.W. (1994) Evidence for Executive Dysfunction in Autism. *Neuropsychologica,* 32, 4, 477-492.

Jazbec, S., Pantelis, C., Robbins, T., Weickert, T., Weinberger, D.R. & Goldberg, T.E. (2007) Intra-dimensional/extra-dimensional set-shifting performance in Schizophrenia: Impact of distractors. *Schizophrenia Research* 89, 339–349.

Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Leitner, Y (2014). The Co-Occurrence of Autism and Attention Deficit Hyperactivity Disorder in Children – What do we know? *Frontiers in Human Neuroscience.* 8, 268.

Leeson, VC, Robbins, TW, Matheson, E, Hutton, SB, Ron, MA, Barnes, TRE & Joyce, EM (2009). Discrimination Learning, Reversal, and Set-Shifting in First-Episode Schizophrenia:

Stability Over Six Years and Specific Associations with Medication Type and Disorganisation Syndrome. *Biol. Psychiatry*, 66(6), 586-593.

Luck, SJ & Gold, JM (2008) The Construct of Attention in Schizophrenia. *Biol Psychiatry*. 2008 July 1; 64(1): 34–39

Luck, S.J., Hahn, B., Leonard, C.J., & Gold, J.M. (2019a). The hyperfocusing hypothesis: A new account of cognitive dysfunction in Schizophrenia. Schizophrenia Bulletin, 45, 991–1000. https://doi.org/10.1093/schbul/sbz063

Luck, S.J., Leonard, C.J., Hahn, B., & Gold, J.M. (2019b). Is selective attention impaired in Schizophrenia? Schizophrenia Bulletin, 45, 1001–1011. https://doi.org/10.1093/schbul/sbz045

de Moivre, A (1738), The Doctrine of Chances: A Method of Calculating the Probabilities of Events in Play. Second Edition.

Ozonoff, S., Cook, I., Coon, H., Dawson, G., Joseph, R.M., Klin, A., McMahon, W.M., Minshew, N., Munson, J.A., Pennington, B.F., Rogers, S.J., Spence, M.A., Tager-Flusberg, H., Volkmar, F.R. & Wrathall, D. (2004). *Journal of Autism and Developmental Disorders.* 34, 2, 139-150.

Pantelis, C., Barber, F.Z., Barnes, T.R., Nelson, H.E., Owen, A.M., Robbins, T.W., (1999). Comparison of set-shifting ability in patients with chronic Schizophrenia and frontal lobe damage. *Schizophr. Res.* 37, 251-270.

The Psychological Corporation (2000). *Wechsler Adult Intelligence Scale III UK edition (WAIS-III UK; 3rd ed.)*. London, UK.

The Psychological Corporation (2008). *Wechsler Adult Intelligence Scale IV UK edition (WAIS-IV UK; 4th ed.)*. London, UK.

van Rooij, I., & Baggio, G. (in press). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science.* https://doi.org/10.31234/osf.io/7qbpr

Shypp, BE & Eimas, PD (1964). Intradimensional and extradimensional shifts in the rat. *J Comp. Physiol. Psychol*. 57, 357.

Smith, H., Carter, A.S., Blaser, E., & Kaldy, Z. (2019). Successful attentional set-shifting in 2-year-olds with and without Autism Spectrum Disorder. *PloS one*, *14*(3), e0213903. https://doi.org/10.1371/journal.pone.0213903

Turner, B.M., & Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. Journal of Mathematical Psychology, 56, 69–85.

Wallsten, T.S., Pleskac, T.J., & Lejuez, C.W. (2005). Modeling behavior in a clinically diagnostic sequential risk-taking task. *Psychological review*, *112*(4), 862–880.

Woodbury-Smith, M.R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. Journal of Autism and Developmental Disorders, 35(3), 331-335.

Yechiam, E., Goodnight, J., Bates, J.E., Busemeyer, J.R., Dodge, K.A., Pettit, G.S., & Newman, J.P. (2006). A formal cognitive model of the go/no-go discrimination task: Evaluation and implications. *Psychological Assessment, 18*(3), 239–249.

Footnotes

[1]Six correct sequential responses is the most common rule, but Hughes et al (1994) require eight, and Pantelis et al (1999) claim to also require eight (ibid, p255), but the reported data is not consistent with this and we assume they also require six.

[2] More precisely, $\frac{2,767,184,147,267}{8,796,093,022,208}$ (de Moivre, 1738). Assuming a criterion of eight correct responses this drops to $\frac{91,877,612,245}{1,099,511,627,776}$, or around 8%.

[3]Another way to see this is to note that whilst AQ alone predicts diagnosis ($p<.001$), adding errors at ED – ID does not add any predictive power, and the weight on ED - ID errors is not significant ($p>.05$).