The Neural Circuitry Mediating Shifts in Behavioral Response and Cognitive Set in Autism

Keith M. Shafritz, Gabriel S. Dichter, Grace T. Baranek, and Aysenil Belger

Background: Recent studies have suggested that the social and cognitive impairments in autism are associated with neural processing deficits in specific brain regions. However, these studies have primarily focused on neural systems responsible for face processing and social behaviors. Although repetitive, stereotyped behaviors are a hallmark of autism, little is known about the neural mechanisms underlying these behaviors in the disorder.

Methods: We used functional magnetic resonance imaging (fMRI) to investigate the neural correlates of shifts in behavioral response and cognitive set in 18 individuals with high-functioning autism and 15 neurotypical control participants. Participants performed a target detection task specifically designed to distinguish shifts in response from shifts in cognitive set.

Results: Individuals with autism showed lower accuracy on response shifting trials, independent of whether those trials also required a shift in cognitive set. Compared with control subjects, participants with autism showed reduced activation in frontal, striatal, and parietal regions during these trials. In addition, within the autism group, the severity of restricted, repetitive behaviors was negatively correlated with activation in anterior cingulate and posterior parietal regions.

Conclusions: These results suggest that executive deficits and, by extension, repetitive behaviors associated with autism might reflect a core dysfunction within the brain’s executive circuitry.

Key Words: Autism, cognitive set, executive function, fMRI

Autistic spectrum disorders (ASD) are characterized by impairments in social interaction, impairments in communication, restricted and repetitive behaviors, and a characteristic course (1). Recent epidemiological data suggest a more widespread occurrence than previously indicated, with a prevalence up to .6% (2–4). Symptoms first appear in early childhood and persist into adulthood, and long-term outcomes include low academic and occupational achievement, with the most severe cases requiring specialized living situations.

Empirical evidence from neuropsychological studies has led to the suggestion that an appropriate model for the repetitive behaviors often observed in autism is a deficit in executive functioning (EF), a collection of mental processes responsible for planning, inhibition of prepotent or inappropriate behavioral responses, and the execution of appropriate responses (5,6). Individuals with autism perform poorly on certain tests of EF, such as the Tower of London (TOL), Wisconsin Card Sorting Test (WCST), and extradimensional shifts in the Cambridge Neuropsychological Test Automated Battery (CANTAB) (5,7–13). Although not all studies have found EF deficits in ASD (13–15), impairments seem to be most profound and universal in the domains of cognitive flexibility, set shifting, and inhibition of prepotent responses (5,7,8,16,17). Poor performance on these measures and related EF tasks has also been observed in individuals with damage to the frontal lobes (18,19) or the basal ganglia (20,21), suggesting that these regions are critical for intact EF abilities.

Functional neuroimaging studies that have investigated the neural circuitry of EF in nonclinical populations have demonstrated the importance of frontal–striatal interactions during EF tasks. Activation in these regions occurs during specific aspects of the WCST (22–24), the Tower of London (TOL) (25,26), and tasks of inhibitory control, such as the go/no-go and stop tasks (27). Recent studies designed to isolate the neural systems mediating specific aspects of EF have shown that certain regions, such as the dorsolateral prefrontal cortex (DLPFC) and striatum, seem intricately involved in some aspects of EF tasks but not others (28–30).

Despite neuropsychological evidence of some impairment in EF abilities in ASD and empirical evidence that certain EF deficits correlate with restricted and repetitive behaviors (16), little is known about the brain mechanisms mediating EF deficits in the disorder. Rather, most neuroimaging studies in autism have focused on social deficits, despite evidence that social deficits might be accompanied by EF deficits in this disorder (31–33).

One would expect that because EF has long been associated with frontal–striatal neural systems, poor EF abilities would be associated with hypoactivation in these regions. However, the few imaging studies that have investigated EF deficits in autism have yielded conflicting results. Whereas one study found that regions within frontal cortex normally engaged by motor sequence learning were not recruited by individuals with autism (34), another study suggested that brain regions mediating inhibitory and set shifting behaviors were overactive in ASD individuals (35). A third study has indicated reduced activity in medial frontal regions during response inhibition (36). Because EF comprises several individual processes, these conflicting results might be due in part to differences in the specific components of EF being tested within each study. Therefore, a primary goal of the present study was to further characterize the neural substrates of specific EF deficits in autism.

The current study used event-related functional magnetic resonance imaging (fMRI) to investigate the neural circuitry of...
two dimensions of EF in individuals with high-functioning autism (HFA). We used a target detection task specifically designed to isolate the neural systems mediating shifts of behavioral responses and shifts of cognitive sets (28). Participants viewed sequences of geometric shapes and were required to alter ongoing responses when a predetermined target shape appeared. Non-target distractor shapes that did not require a change in behavioral response were interspersed within the sequence of shapes. The shape identified as the target changed periodically throughout the study, resulting in two conditions: a shift in behavioral response associated with the presentation of a target, and a shift in the rules governing stimulus-response associations. We hypothesized that because EF deficits are commonly observed in autism, HFA participants would perform poorly on the task compared with control participants. Specifically, we predicted that HFA participants would be less accurate than control participants in altering their ongoing response when a target shape was presented. We also predicted that HFA participants would show the lowest accuracy on trials requiring a shift in both behavioral response and cognitive set (‘target-shift’ trials), because these trials have the highest cognitive demand. In addition, we predicted that HFA participants would show reduced activation in frontal, striatal, and parietal regions relative to control participants for trials requiring a shift in behavioral response.

Methods and Materials

Subjects

Eighteen individuals with HFA (16 male, 2 female; mean age 22.3 years, SD = 8.7) and 15 age- and IQ-matched comparison subjects (13 male, 2 female; mean age 24.3 years, SD = 6.2) with no history of psychiatric or neurological disorder participated in this study. Participants in the autism group were recruited through the North Carolina Neurodevelopmental Disorders Research Center Subject Registry and the Treatment and Education of Autistic and Related Communication Handicapped Children program. The DSM-IV diagnoses of autism were based on clinical interviews with the Autism Diagnostic Interview—Revised (ADI-R) (37) and the Autism Diagnostic Observation Schedule (ADOS) (38). The IQ scores were obtained with the Wechsler Abbreviated Scale of Intelligence (WASI) (39). The average (SD) IQ scores for the autism group were: Verbal = 103.3 (19.3), Performance = 101.1 (16.1), Full-Scale = 102.5 (17.6). One HFA participant whose data appear in the final data analysis was taking psychotropic medication (Adderall 10 mg/day).

Participants in the control group were recruited through advertisements in the local community and were screened through detailed interviews to ensure absence of psychiatric, neurological, or developmental disorders. The average IQ scores for the control group were: Verbal = 111.2 (17.1), Performance = 109.3 (11.0), Full-Scale = 111.4 (15.1). No statistical differences were observed between the two groups for either age or IQ. Owing to excessive motion in the MRI scanner, three participants in the HFA group and one participant in the control group were removed from the final data analysis. The two groups were still matched for age and IQ after removal of these participants.

All participants (and guardians, if applicable) received a complete verbal description of the study and provided written informed consent, approved by the University of North Carolina School of Medicine Committee on the Protection of the Rights of Human Subjects and the Duke University Health System’s Institutional Review Board.

fMRI Task

Participants performed a target detection task (28) during which geometric shapes (squares, triangles, or circles) were presented one at a time. Participants were required to classify each stimulus as a “target” or “non-target” on the basis of its shape and respond with an appropriate button press. Participants responded to each shape presented, pressing one button for all non-targets and an alternate button for targets. Stimuli consisted of geometric shapes of varying size and color. Frequently (94%) occurring squares (“standards”) and infrequently occurring circles (3%) and triangles (3%) were presented centrally against a white background for 500 msec, with a stimulus onset asynchrony of 1500 msec. The task was presented using CIGAL software (40) as an event-related design, with 5 target and 5 non-target novel stimuli embedded within 154 standard stimuli in each imaging run (11 runs total). Each event (presentation of a target or non-target novel stimulus) was separated by a minimum of nine standard stimuli (i.e., 15 sec) to adequately observe the hemodynamic response for each event and to ensure that event-related responses did not overlap.

An instruction screen was presented (5 sec) at the beginning of each run, designating either circles (“Targets = •”) or triangles (“Targets = ▲”) as the target stimulus for that run. The stimulus designated as the target changed after every two runs, resulting in runs in which the target shape changed at the onset of the run (“shift runs”) or was maintained from the previous run (“maintain runs”). Participants were instructed to respond to every stimulus presented and to press one button for all non-target stimuli (including standards) and an alternate button for target stimuli (Figure 1). In this manner, motor activity related to making a button-press response was incorporated into the baseline activation for our task (see the following functional image analysis).

Periodically altering the target stimulus produced two discernable types of target trials. Targets occurring during maintain runs required a shift in behavioral response only (behavioral response shifts), whereas those occurring during shift runs required shifts in both behavioral response and the cognitive mapping of stimulus-response associations (cognitive set shifts). Therefore,
the task distinguished between behavioral response shifting (altering the ongoing button press response when a target stimulus occurred) and cognitive set shifting (remapping the target stimulus during shift runs). Imaging session parameters can be found in Supplement 1.

Functional Image Analysis

Before statistical analysis, head motion was analyzed by center of mass measurements in three orthogonal planes. Imaging sessions with motion > 1 voxel in any of the three dimensions resulted in rejection due to motion artifact. With these criteria, data from three HFA participants and one control participant were removed from further analysis.

Images were then temporally realigned, motion corrected, normalized, and Gaussian-filtered (full-width at half-maximum = 8.0 mm) with the SPM99 program (Wellcome Department of Cognitive Neurology, London). With in-house software, individual subject t-maps were generated for each stimulus type (targets, novels, targets-shift, targets-maintain, novels-shift, novels-maintain) by identifying voxels whose average activity correlated with an empirically derived hemodynamic response template from a nonclinical population (41). For the purposes of individual subject t-map creation, the five images preceding each event (presentation of a triangle or circle) served as the baseline for each event's hemodynamic response. This baseline controlled for neural activity associated with simple button-press responses accompanying the presentation of the standard squares. Image time segments from similar events (i.e., targets or non-target novels) were then averaged, and for each stimulus type, the mean of the five pre-event images was subtracted from each of the 15 images comprising the averaged segment to generate baseline-adjusted images. The resulting maps from all participants in the HFA group and all participants in the control group were averaged and these group-averaged t-maps were displayed on group composite anatomical images with a statistical threshold of \( p < .001 \), adjusted for multiple comparisons with a cluster-filter of 10 contiguous pixels (42).

To directly compare activations between the two diagnostic groups, between-group contrast maps were created with a random-effects model. For each task component, a \( t \) test compared activation at each voxel between the diagnostic groups. The resulting maps from each comparison were displayed on a composite anatomical image with a statistical threshold of \( p < .001 \), cluster-filtered to 10 contiguous pixels.

Finally, because EF deficits might correlate with restricted and repetitive behaviors in autism (16), an exploratory region of interest (ROI) analysis examined the relationship between blood oxygen level dependent (BOLD) signal change to target stimuli and the severity of restricted, repetitive behaviors (RRB) for the HFA participants. Regions were selected on the basis of imaging results from a previous study using this task (28) and consisted of the DLPC (Brodman area [BA] 9 and 46), ventrolateral prefrontal cortex (VLPFC), the anterior cingulate cortex (ACC; BA 24 and 32), the intraparietal sulcus (IPS) within posterior parietal cortex (PPC), and the basal ganglia (BG). Functional ROIs were demarcated on the basis of the results of the whole-brain voxelwise analysis described earlier. The average percent signal change from baseline in each of these regions for each stimulus type was correlated with individual algorithm scores for the Restricted, Repetitive, and Stereotyped Patterns of Behaviour Domain (RRB) of the ADI-R. A similar analytical technique correlated BOLD signal change in each ROI with task performance.

---

**Figure 2.** Performance data for the functional magnetic resonance imaging (fMRI) task. Accuracy and reaction time data were collected during the fMRI sessions for all trial types. Data are sorted by clinical group and then by trial type. All target, novel, and standard trials are shown first, followed by target trials divided into shift and maintain runs, then novel trials divided into shift and maintain runs. *High-functioning autism (HFA) participants performed more poorly than control subjects on target trials, as indicated by lower accuracy for all targets, targets-shift, and targets-maintain.

---

**Results**

**Behavioral Performance**

Accuracy (percent correct) and reaction time (RT) data are shown in Figure 2. Statistical analyses of RT data were performed only on correct trials in order to limit the effect of simple RT/accuracy trade-offs. Two separate Group × Trial Type analyses of variance (ANOVAs), one for accuracy and one for reaction time, compared performance for target, novel, and standard trials among the HFA and control groups. Performance differences were found between the trial types, consistent with previous findings using this task with a group of typically developing individuals (28). For accuracy, there was a main effect for Trial Type \( F(2,54) = 94.09, p < .001 \), a main effect for Group \( F(1,27) = 8.88, p < .01 \), and a Group × Trial Type interaction \( F(2,54) = 4.82, p < .05 \). For RT, there was a main effect for Trial Type \( F(2,54) = 105.71, p < .001 \) and a Group × Trial Type interaction \( F(2,54) = 4.01, p < .05 \) but no significant main effect for Group \( F(1,27) = 1.11, p = ns \).

To further examine the effects of trial type within each group, separate within-group ANOVAs compared performance for target, novel, and standard trials in the HFA and control groups.
Both HFA and control participants showed accuracy differences between the task conditions \(F(2,28) = 53.90, p < .001\) for HFA group; \(F(2,26) = 44.42, p < .001\) for control group as well as reaction time differences \(F(2,28) = 48.39, p < .001\) for HFA group; \(F(2,26) = 76.66, p < .001\) for control group. Post hoc paired samples \(t\) tests then revealed that for both HFA and control participants, accuracy was lower and RT was longer for target trials compared with both novel trials and standard trials \((p < .001\) for all comparisons). In addition, lower accuracy and longer RTs were observed for novel trials compared with standard trials \((p < .05\) for all comparisons). For both HFA and control participants, no significant differences in accuracy or RT were observed when comparing target and novel trials during shift runs with those trials during maintain runs.

Next, \(t\) tests compared performance between the HFA and control participants for each task condition. For target trials, the HFA group showed lower accuracy compared with the control group \(t(27) = 2.68, p < .05\) but no differences in RT. Dividing target trials into target-shift and target-maintain trials, HFA participants showed lower accuracy for both target-shift \(t(27) = 2.62, p < .05\) and target-maintain \(t(27) = 2.37, p < .05\) trials, relative to control participants. No between-group performance differences were observed for either novel or standard trials.

**Imaging Data**

For control participants, within-group average activations in response to target and novel trials showed that target trials engaged the DLPFC (BA 9 and 46), premotor cortex (PMC; BA 6), a portion of ACC (BA 24 and 32), PPC in the IPS, thalamus, BG, and cerebellum, but novel trials did not strongly recruit these regions (Figure 3A). In addition, the VLPFC and anterior insular cortex were recruited by both target and novel trials, but the degree and extent of activation were greater for target than for novel trials. For HFA participants, the same analysis showed that the regions activated by control participants were either not significantly recruited or seemed to be equally engaged for target and novel trials, with the exception of VLPFC/anterior insula (Figure 3B).

Comparing activation for target trials between the diagnostic groups revealed that the control group activated the DLPFC, IPS, and BG to a greater degree than did the HFA group (Figure 4). When target trials were divided into target-shift and target-maintain trials and activations were compared between the two diagnostic groups, a similar pattern was observed (Figure 5). For target-shift trials, the control group recruited the DLPFC, ACC, IPS, and BG to a greater degree than did the HFA group (Figure 5A). For target-maintain trials, the control group activated the DLPFC, PMC, IPS, and BG to a greater degree than did the HFA group (Figure 5B). Table 1 lists Talairach coordinates for these activation differences. Novel trials did not elicit significant between-group activation differences, except for a small region in the left insular cortex (Talairach coordinates for center of activation: \(x = -40, y = -11, z = +18\)). In addition, there were no regions in which the HFA group showed greater activation than the comparison group for any of the task components.

**Correlation Between Imaging and Clinical Data**

Analyses of relations between BOLD activation and RRB scores revealed significant negative correlations between ADI-R...
No other correlations higher ADI-R RRB scores showed the smallest BOLD responses accuracy for target stimuli. The threshold for significant activation was $p < .001$ (uncorrected), with a cluster-filter of 10 contiguous pixels. Cluster sizes are uninterpolated values, with the extent of activation relating to functional voxels, not anatomical voxels.

Target trial BOLD scores and activation to target stimuli in ACC ($r = - .55$, $p < .05$) and left IPS ($r = - .56$, $p < .01$), indicating that subjects with higher ADI-R RRB scores showed the smallest BOLD responses to target events in these regions (Figure 6). No other correlations were found to be significant. In addition, BOLD activation to target stimuli was not significantly correlated with response accuracy for target stimuli.

**Discussion**

Individuals with HFA failed to recruit neural circuitry typically engaged during the performance of an EF task designed to differentiate shifts in behavioral response from shifts in cognitive set. Specifically, we found that during response shift trials, control participants recruited a neural system comprising the DLPFC, ACC, premotor cortex, IPS, BG, thalamus, cerebellum, and VLPFC/anterior insula. Task-related activation in the autism group, however, was limited to the VLPFC (Figure 3). Direct statistical comparison of BOLD activation between control and HFA participants showed greater activation to target trials for the control group in DLPFC, BG, and IPS (Figure 4). This pattern of reduced activation in the autism group in frontal, striatal, and parietal regions was observed when target stimuli were subdivided into target-shift and target-maintain trials (Figure 5), indicating that these functional deficits occurred for all trials requiring a response shift and not only during runs requiring a shift in cognitive set. Performance data indicated that HFA participants were less accurate than control participants in responding to target trials. Although performance on the task differed between groups, task performance was not correlated with brain activation in any a priori ROI. Therefore, these activation differences cannot be accounted for solely by performance differences; rather, the differences likely reflect neural dysfunction in HFA. We suggest that the reduced activation among HFA participants signifies a global neural processing deficit related to the disorder and might indicate an ineffective neural strategy for altering ongoing behavioral responses.

The current findings are important, given the empirical evidence suggesting that cognitive flexibility and other EF deficits are apparent in autism (5,43) and that some EF deficits might be more severe in autism than in other childhood disorders (7). These findings complement and extend recent imaging studies of the neural substrates of EF deficits in autism, such as reports of reduced PFC activation during motor sequence learning (34), auditory target detection (44), and response inhibition (36). The task used in the current study isolated two components of EF often measured through the WCST (i.e., the shifting of behavioral responses and cognitive sets) (28). Because the WCST is complicated by its vague instructions and complex design, we sought to simplify the behaviors associated with this task and make it easier for participants to understand the task instructions. Despite the simplification, our task measures at least three properties of EF and cognitive flexibility tested in the WCST, specifically the maintenance of a stimulus-response rule, the alteration of ongoing prepotent responses, and the periodic shifting of stimulus-response rules. Therefore, our findings are consistent with current views of executive dysfunction in autism as driven primarily by deficits in cognitive flexibility (17) and the ability to inhibit prepotent ongoing responses (5).

An intriguing finding in the current study was that activation to target stimuli within ACC and IPS declined as a function of RRB severity as measured by the ADI-R (Figure 6). This negative relationship between RRB scores and BOLD activation in brain regions associated with executive and attentional processes (45–47) might indicate that repetitive behaviors in autism are driven, at least in part, by more general deficits in attention, target detection, and response preparation (48). Because the correlation results were part of an exploratory analysis, they should be considered preliminary; further investigation into the possible relations between neural activation during cognitive tasks and RRB severity in autism is necessary to draw more definitive conclusions. Despite their preliminary nature, these findings, together with the hypoactivation observed in the autism group, fit...
well within models suggesting that the restricted and repetitive behaviors commonly observed in autism might be related to deficits in certain domains of EF, such as cognitive flexibility and response inhibition (6,16).

The results of the current study indicate that deficits in cognitive flexibility observed in the HFA group were limited to shifts of behavioral response, with cognitive set shifting ability unaffected by the disorder, at least when instructions to shift cognitive set are overt. We predicted that HFA participants would have the greatest difficulty with target-shift trials, because these trials require shifts in both behavioral response and cognitive set. To successfully respond to these trials, participants were required to alter the stimulus–response associations governing the previous run. For example, if the previous run required shifting responses for triangles but not for circles, the participant must now shift responses for circles and continue responding normally for triangles. Therefore, during shift runs, participants were required to generate shifts in responses for target stimuli in addition to creating shifts in cognitive set. Because there were no performance differences between the target-shift and target-maintain trials for HFA participants, we suggest that HFA individuals demonstrate an intact ability to shift cognitive set when the cognitive shift rule is overtly displayed.

Performance data for non-target novel trials provide corroborating evidence for an intact cognitive set shifting ability during our task in HFA individuals. These data indicated that accuracy for novel trials within the HFA group was not significantly different from the control group, even for novel trials during shift runs. If overt cognitive shifts were affected in autism, we would have observed poor performance in the HFA group during novel-shift trials in addition to target-shift trials. Although HFA individuals might be prone to fall into repetitive behavior patterns and thus would be expected to have the greatest difficulty with target-shift and novel-shift trials, this behavioral pattern was not observed in our data. Therefore, we conclude that the participants with autism did not have difficulty understanding that the shape previously identified as the target was now a non-target novel stimulus.

Our findings of reduced activation in multiple brain regions during set shifting are consistent with at least three other reports of hypoactivation in frontal regions during EF-related tasks (34,36,44). However, Schmitz et al. (35) reported significantly increased brain activation in multiple areas during a motor inhibition task, a cognitive interference task, and a set shifting task. Most relevant in the present context is the hyperactivation in the HFA group in response to set shifting reported by this research group. It should be noted that the set shifting task employed by this group required a shift of both behavioral response and cognitive set on all switch trials, whereas the current paradigm differentiated behavioral and cognitive set shifting. Nonetheless, even on trials requiring both behavioral and cognitive set shifts, we observed hypo- rather than hyperactivation in the HFA group. Moreover, our current finding of selective set shifting deficits in autism might explain why these behaviors seemed to be preserved in HFA individuals, as reported by the other research group (35). Further imaging studies using paradigms that tap specific components of EF (e.g., by isolating the inhibition of prepotent responses from the execution of appropriate responses) are necessary to shed light on these contradictory findings.

In summary, our findings demonstrate neural hypoactivation during an EF task in HFA. Differences in brain activation between HFA and control participants were apparent both for between-


