VISUAL COGNITION, 2001, *8* (1), 67–101

A test of central coherence theory: Can adults with high-functioning autism or Asperger syndrome integrate objects in context?

Therese Jolliffe and Simon Baron-Cohen

Departments of Experimental Psychology and Psychiatry, University of Cambridge, UK

Weak central coherence was investigated by exploring the conceptual integration of objects. Normally intelligent adults with either autism or Asperger syndrome were given two novel experiments. Experiment 1, the Object Integration test, had sets of line drawings depicting objects and people. Each set had to be either visually integrated to make the most coherent scene, or compared for similarities. The clinical groups were significantly impaired in their ability to integrate objects, but they were not impaired in looking for similarities. Experiment 2, the Scenic test, presented black line drawings of scenes containing an item that was inappropriate for the context. Participants were required to describe the scenes, identify the type of scene and context-inappropriate object, and locate a name (incongruent) object as quickly as they could. The clinical groups' descriptions suggest that they did not spontaneously pay preferential treatment to local details, nor were they faster at locating a named incongruent object. Whereas only a few of their descriptions lacked coherence, there was a deficit in both their ability to spontaneously notice and identify incongruent objects, as well as to identify the scenes. These tests provide support for Frith's (1989) central coherence hypothesis. Conceptual or high-level processing seemed inferior, whereas perceptual or low-level processing seemed normal, but not superior. Poor performance on these tasks characterized the majority of clinical participants, but those with autism performed at a consistently poorer level than those with Asperger syndrome. Possible explanations for the clinical groups' difficulties are explored along with suggestions for future research.

Please address all correspondence to T. Jolliffe, Dept. of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, UK.

We are grateful to a large number of people for helping us recruit subjects: Lorna Wing, Wendy Phillips, the National Autistic Society, Pat Howlin, Clive Robinson, Ben Sacks, and Pam Yates. We are also grateful to Simon and Denise Jolliffe who designed and drew most of the stimuli. The first author was supported by an MRC Studentship during the period of this work. The Harold Hyam Wingate Foundation also provided her with valuable financial support. Both studies reported here were submitted in part fulfilment of her PhD degree at the University of Cambridge.

http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280042000036

Frith (1989) noted that normal individuals integrate information for higher-level meaning. She also noted that this ability is diminished in autism, which gives rise to an imbalance in the normal global–local dichotomy, akin to a deviant processing style. Thus, Frith suggests a problem in integrating information for meaning with the result that there is a tendency to process locally. She called this processing characteristic "weak central coherence".

The laboratory evidence for a difficulty in integrating information for meaning, tends to be linguistic in nature (Happé, 1997; Jolliffe & Baron-Cohen, 1999b, c). The evidence for a difficulty in integrating information in the visual modality seems to be based at the perceptual rather than conceptual level, which has been referred to as low-level integration (Happé, 1996). Thus, Happé demonstrated that children with autism are less likely to succumb to visual illusions due to a tendency not to integrate elements of a stimulus. Similarly, Mottron and Belleville (1993) found that a normally intelligent adult with autism experienced difficulties in perceiving the geometric "impossibility" of figures (for example, the Penrose triangle) which was interpreted as a defect in relating (integrating) parts of the figure. Not surprisingly, some children and adults with autism have an enhanced ability to process locally or rapidly detect embedded figures on the Embedded Figures test (EFT: Witkin, Oltman, Raskin, & Karp, 1971; see Jolliffe & Baron-Cohen, 1997; Shah & Frith, 1983), whereas Mottron, Belleville, and Ménard (1996) found that a normally intelligent group of individuals with autism tended to draw more local and less global features of objects.

There is only one visual study which has addressed the central coherence hypothesis at the visuo-conceptual level (Jolliffe & Baron-Cohen, 1999a). This paper presented a modified version of the Hooper Visual Organization test (Hooper, 1983) to normally intelligent adults with either autism or Asperger syndrome. This test presented line drawings depicting simple objects that had been cut into pieces or fragments and arranged in a puzzle like fashion. The participant was required to mentally (conceptually) integrate the fragments in order to identify what each object would be if the fragments were put together correctly. Individuals with an autism spectrum condition were significantly impaired in their ability to integrate fragments holistically. This led the authors to conclude that there was an autistic spectrum deficit in the conceptual integration of visual elements.

The evidence from studies of autism seem to suggest a difficulty in integrating information, be it visual or linguistic in nature. This is in accord with Frith's weak central coherence hypothesis and leads to two key questions. First, whether the findings reflect a mere (non-conscious) processing preference or are part of a much more serious deficiency in integrating information. This is an open question and one which will be addressed in this paper. Second, the evidence suggests that there might be some unique relation between the global and local levels, so that impaired processing at one level results in enhanced processing at the other level. It could also be that the two levels are actually independent of one another, but suffer from universally weak central coherence in individuals with autism. This paper seeks to address these issues with two studies, both of which look at visual processing problems in terms of difficulties in conceptual integration.

EXPERIMENT 1
THE OBJECT INTEGRATION TEST

Introduction

The Object Integration test is a novel test designed to examine visuo-conceptual integration. Thus this test has sets of five meaningful stimuli which when visually integrated make up a coherent scene (or context). The stimuli consist of line drawings which depict objects and people. These are arranged five to a page, so that there is also one or two people[1] and always three or four objects. The participant's job is to visually integrate these stimuli so as to identify the most coherent scene, and to do this they are required to select which one of the five objects is incongruent or odd for the established context. It is the need to identify what is incongruent from amongst the objects which enforces visuo-conceptual integration. Though all five single representations can be visually integrated to construct a scene, only four out of the five make a visually coherent scene.

There is also a control task, which is the "category" condition. In this condition there are always five objects and all need to be compared to identify the similarities between the items so as to again determine which item is odd. This control task requires some degree of coherence since one's stored knowledge of the items needs to be brought together so that the odd item can be detected. It is suspected that our normally intelligent clinical groups would have a sufficient degree of coherence to enable them to cope with this task. This is because children with autism have been found to have categorization skills consistent with their mental age (Ungerer & Sigman, 1987) and because the requirement of the category task is very similar to the Wechsler (1949, 1955, 1974, 1981) Similarities subtest, which, although linguistic, finds individuals with an autistic disorder to be unimpaired. Although we suggest that the clinical groups have sufficient coherence to perform normally on the Category task, evidence from perceptual studies such as work on visual illusions (Happé, 1996) suggests that coherence appears to be extremely weak. However, it is likely that on the visual illusions task as with many other perceptual tasks the difference between the autism group and controls is more likely to be due to the controls failing to resist a gestalt. This would suggest that tasks that are perceptual may not suffer from

[1] For convenience the people will be referred to as objects.

particularly weak central coherence compared with tasks that are conceptual. Comparing coherence at the perceptual and conceptual levels is something that will be addressed in this paper.

As discussed, unless coherence is extremely weak the central coherence hypothesis would not predict a problem on the Category task. This is because the Category task requires a bottom-up approach consisting of comparing objects with their stored knowledge of the item. This differs from the Integration test which is top-down, since it is only when the objects are grouped together for maximum coherence that the scene and hence incongruent object can be identified. If individuals with their autism or Asperger syndrome have weak central coherence, the prediction is that they will be less proficient at mentally synthesizing the objects so as to construct the most coherent context.

Participants

Fifty-one adults participated in the experiment. These comprised 17 with high-functioning autism, 17 with Asperger syndrome, and 17 normal adult control participants. The normal adults acted as a comparison group for the two clinical groups. The majority of clinical participants were tested in their place of residence, except where some preferred to be tested at the university. All control participants were tested in a quiet room in the university.

The histories of all 17 of the adults with autism were those of classical autism, i.e., autism accompanied by language delay[2] and therefore met established diagnostic criteria (DSM-IV: American Psychiatric Association, 1994). The symptoms of the adults with autism had lessened considerably and these adults would be considered to be in the "residual" category according to DSM-III-R (American Psychiatric Association, 1987).

Some of the 17 individuals with Asperger syndrome had at one stage been diagnosed as having autism but their diagnosis was amended to that of Asperger syndrome in that they exhibited symptoms of autism without a clinically significant delay in early language development and hence met the criteria for Asperger syndrome when these were established by ICD-10 (World Health Organization, 1992). Some of the adults with Asperger syndrome had been diagnosed in childhood or adolescence, but many had not received their diagnosis until adult life.

The 17 control participants were taken from the general population of Cambridge. All groups of participants were briefly screened on their medical and family history. All were medication-free at the time of testing, and there was no known history of neurological disorder or a head injury. The normal control participants did not have a family history of autism of Asperger syndrome and

[2]Language delay is defined quite narrowly: not using single words by 2 years, and not using communicative phrases by 3 years. Pragmatics is not included.

all were free from any history of psychiatric disorder. The 17 control participants were chosen to match the clinical groups as closely as possible with respect to the characteristics of age, IQ, sex, and handedness.

All participants were required to be of at least normal intelligence (i.e., scoring ≥85) on the WAIS-R (Wechsler, 1981, full scale, performance and verbal IQ) and the clinical groups were selected on the basis of their ability to pass both first- and second-order false belief tasks.[3]

Participants were matched on age and IQ (see Table 1). Table 1 gives the participant details of chronological age (CA), verbal IQ (VIQ), performance IQ (PIQ), and full-scale IQ (FSIQ). Four one-way ANOVAs revealed no significant differences between groups on any of these variables: CA $F(2, 48) = 0.59$, $p = .56$; VIQ $F(2, 48) = 0.51$, $p = .60$; PIQ $F(2, 48) = 0.58$, $p = .57$; and FSIQ $F(2, 48) = 0.10$, $p = .91$. The sex ratio in all three groups was 15:2 (m:f), reflecting the sex ratio found in these clinical groups in other studies (Klin, Volkmar, Sparrow, Cicchetti, & Rourke, 1995; Wing, 1981). The sex ratio was the same across groups, because this study has a spatial element to it and aspects of spatial ability are known to be superior in males (Maccoby & Jacklin, 1975; Voyer, Voyer, & Bryden, 1995). The groups were closely matched on handedness, there being 15 right-handed and 2 left-handed individuals in the normal and high-functioning autism group, and there being 14 right-handed and 3 left-

TABLE 1
Participant characteristics

Participant group	*CA*	*VIQ*	*PIQ*	*FSIQ*
Normal (n = 17)				
Mean	30.00	106.47	105.24	106.35
SD	9.12	10.94	14.00	12.72
Range	(18–49)	(87–127)	(85–134)	(88–133)
Autism (n = 17)				
Mean	30.71	107.59	101.77	105.12
SD	7.84	14.37	13.06	13.47
Range	(19–46)	(88–135)	(85–132)	(90–133)
Asperger (n = 17)				
Mean	27.77	110.82	100.29	107.12
SD	7.81	13.51	14.23	14.34
Range	(18–49)	(89–130)	(85–133)	(86–132)

[3]Participants were given first- and second-order theory of mind tests. The first-order task was a version of Perner's (Perner, Frith, Leslie, & Leekam, 1989) Smarties task. The second-order task was Baron-Cohen's (1989a) ice-cream van test. Whereas all participants passed the first-order task, 5 out of 51 participants failed the second-order task. These included one participant with Asperger syndrome, two with high-functioning autism, and two normal control participants. These participants were retested on a new variation of the second-order belief task and all were found to pass.

handed in the Asperger group. All participants were born in England and English was their first language. All three groups contained participants from various socio-economic backgrounds and the three groups were broadly equivalent in terms of educational attainment. Several individuals within each group were either studying for or held formal qualifications such as a university degree or diploma.

Materials

The stimuli presented consisted of black line drawings of objects and people, some of which were taken from the Snodgrass and Vanderwart (1980) set and some from the British Picture Vocabulary Scale (BPVS: Dunn, Dunn, & Whetton, 1982). These line drawings were presented on sheets of plain white A4-sized paper, so that there was always five black line drawings appearing on a single sheet. There was one set of 10 sheets which formed the object integration condition, and there was one set of 5 sheets which formed the control or category condition, and each of the two test sets had a corresponding set of 2 trial pages. In the integration condition there was on each sheet of paper a combination of objects and people, so that there was always one or two people and always three or four objects. In the category or control task there were five objects. All objects and people appeared in the upright mode and one object (or person) was always placed centrally, whilst the other four were placed in each quadrant. For both of the conditions, the position of the odd object varied and was completely orthogonal, so that each odd item appeared in each of the five possible positions twice for the integration task and once for the category task.

A stopwatch was used in order to time how long it took the participant to determine what the odd object or person was. In order to assist in the recording of response times, there was a large foolscap-size blank white card, the purpose of which was to cover the test pages. This card could be rapidly removed and allowed greater accuracy in recording response times than could be achieved solely by turning over the test pages and starting the stopwatch.

The stimuli that were used in this experiment were presented to 12 individuals in a pilot study. It was found that the pictures on each page were relatively easy to name and relatively unambiguous. However, one pilot participant failed to correctly name one of the objects (the hill) in the integration condition. (The stimuli used in this task can be obtained from the first author.)

Procedure

The set of stimuli was placed in a pile face up on the table, and directly in front of the individual, and on top of this pile of laminated pages was the large foolscap card, which was removed exactly when the timing of participants' responses began.

The trial and test items. The participants were told that they were going to see some pictures. The experimenter removed the covering card and turned to the first trial item of the integration condition. The experimenter drew the participant's attention to the fact that there were five pictures on the page. She named each in turn and stated that four out of the five pictures made up a larger picture or a scene. They were told that this particular scene was about an old lady raking leaves to put into a dustbin. They were also reminded that there was a picture of a pram on the page, and they were told that the pram did not quite fit the scene. All participants could acknowledge that the pram did not fit the scene, so the experimenter turned to the next page of the trial items, which portrayed a circus scene. Participants were encouraged to identify what the odd object or person was in this scene. Given the simplicity of this scene no individual had any difficulty in identifying the incongruent item. The trial items were always presented in the same fixed order and participants had no difficulty with these simple scenes or in understanding the requirements of the task.

The participants were then told that they were going to start the more difficult test items. They were instructed that the procedure was identical, except that this time they were going to be timed to see how long it took them to find what the odd object, or person, was. They were also told to point to or say what the object, or who the odd person, was. They were also instructed to make a guess if they really did not know what was odd. Also they were told that they could not amend their choice as this would distort their response time. Unlike the trial items the test items were presented in a different random order to each participant, in order to avoid order effects.

Upon completion of the integration condition, participants were given the category or control condition. The category condition was always presented after the integration condition. This allowed a check for fatigue effects. Like the previous condition, the trial pages were presented in a fixed order. With the first trial item it was again pointed out that there were five small pictures on each page. These were named in turn by the experimenter and it was mentioned how this time the pictures did not make up a scene, but rather they all formed a category. It was pointed out again that one of the things on the page was odd, and that they were to go through the pages pointing to or saying what was the odd one out. Participants were given the second trial item to work through on their own. Then they continued straight on to the test items as no individual had any difficulty understanding the requirements of this condition. As with the previous condition, the test pages were presented in a different random order, so that there would again be no effect of order. Participants were again reminded that they could not change their choice of item because their response times would be distorted.

It was possible that participants could fail the test due to an inability to recognize one or more of the objects on each page. To examine whether this was the case the experimenter at the end of the test got individuals to name the five objects on each of the pages where they had made a wrong choice.

Scoring. There were two types of performance measure for each condition: (1) the response time, this was the time taken to determine what the odd object or person was; and (2) the accuracy score, this was the odd objects or people correctly identified.

A stopwatch was used in order to time how long it took the participant to make their choice of what object (or person) was odd. The recording of response times was achieved as follows: The stopwatch was started as soon as the card was removed from the top of the pile of stimuli. As soon as the participant pointed to or reported what the odd object was, the experimenter stopped the stopwatch and recorded what the time was.

The test page remained in view until the participant had completed their responses. Participants were given an unlimited amount of time to try to identify the odd object/person. This was to avoid any effects being due to poor attention and memory deficits.

Results

Identifying the incongruent object. Due to the small number of stimuli an item analysis was performed. The three groups scores for each stimulus item is illustrated in Table 2. Participants agreed on what were the more easy or difficult items (Kendall's co-efficient of concordance: integration task, $W = .85, p < .01$; category task, $W = .78, p < .05$). Despite this agreement Table 2 suggests that the clinical groups had a problem with detecting the odd object in about half of the integration scenes, although it seems that they did not actually have a problem in detecting incongruent objects in general. A series of 3×2 Pearson chi-square tests were conducted to determine whether the groups differed on individual stimuli. For the integration condition the groups differed on four stimuli, χ^2 (1): surveying a valley = 6.21, digging for treasure = 6.14, knocking cans off fence = 7.83, busking = 15.53; $p < .05$; all other stimuli were non-significant ($p \geq .12$). The source of the differences were investigated with 2×2 chi-squared tests. The autism group differed from the normal group on all these stimuli, χ^2 (1): surveying a valley = 6.10, digging for treasure = 5.85, knocking cans off fence = 5.76, busking (Yates Continuity Correction to correct for expected frequencies < 5) = 9.67; $p < .05$. The Asperger group differed from the normal group on just one stimulus, knocking cans off fence χ^2 (1) = 5.76, $p < .05$. The two clinical groups differed on one item, busking χ^2 (1) = 6.58, $p < .05$. There were no differences on the category condition ($p \geq .10$).

It is important to determine whether the clinical groups' weaker ability in integrating objects is determined by just a few individuals in each group or characterized the majority. To this end the control group's mean score across

TABLE 2
Number of participants correctly identifying the incongruent object

Subject group (n = 17)	*Incongruent Object*		
Stories	*Normal*	*Autistic*	*Asperger*
Integration condition			
Window cleaning	16	16	17
Getting a taxi	17	16	17
Back packing	17	15	17
Watering a horse	17	15	15
Feeding a giraffe	9	8	10
Reaching an item	15	10	12
Surveying a valley	14	7	11
Busking	17	8	15
Digging for treasure	13	6	8
Knocking cans off fence	12	5	5
Category condition			
Furniture	17	17	17
Toys	17	17	17
Face parts	15	16	16
Vehicles	16	15	17
Stringed instruments	15	10	14

all integration stimuli was calculated. Then the number of participants in this group scoring above (and below) this mean was calculated. This was compared to the number of participants in the clinical groups scoring above (and below) this mean. The analysis revealed that the clinical groups differed significantly from the normal control group, χ^2 aut. (1) = 10.09, $p < .01$; χ^2 Asp. (1) = 4.25, $p < .05$.

Response time. The mean response times for the three groups are illustrated in Table 3. A two-factor repeated measures ANOVA was performed on the mean response time data for the three groups. This ANOVA had a between-participant variable of group, and a within participant variable of condition (integration and category). The ANOVA revealed a significant main effect of group, $F(2, 48) = 10.53$, $p = .001$, and a significant main effect of condition, $F(1, 48) = 285.80$, $p < .001$. The higher-order interaction of group by condition was also significant, $F(2, 48) = 8.49$, $p = .001$.

Given that the group effect was significant, then the group by condition interaction was investigated further to see if there were different group effects for the two conditions. Simple effects showed the effect of group to be significant only for the integration condition, F Integration $(2, 48) = 13.97$, $p < .001$; F Category $(2, 48) = 0.24$, $p = .79$. Since predictions were not made about

TABLE 3
Mean response time for object choice

Participant group (n = 17)	*Integration*	*Category*
Normal		
Mean	11.48	7.02
SD	3.78	2.76
Range	(5.5–18.8)	(3.5–12.2)
Autism		
Mean	19.30	7.94
SD	4.56	3.10
Range	(11.0–26.6)	(3.2–12.6)
Asperger		
Mean	17.43	7.58
SD	3.38	2.43
Range	(10.1–24.0)	(4.7–12.0)

participants' response times, the source of the interaction effect was investigated further using post hoc Newman-Keuls tests. Post hoc comparisons revealed that response times on the integration condition were for the autism and Asperger groups significantly different to that of the normal control group ($p < .01$), although the two clinical groups did not themselves differ ($p > .05$).

Given that the condition effect was significant, it was interesting to investigate whether there were different condition effects for the three groups. Simple effects showed the effect of condition to be significant for all three groups, F aut. $(1, 48) = 168.05, p < .001$; F Asp. $(1, 48) = 126.14, p < .001$; F norm. $(1, 48) = 25.27, p < .001$.

Object naming and omissions. Since it was the integration condition that was difficult for the clinical groups, the remaining analysis shall focus on this condition. To see whether a problem in object recognition could underlie poor performance, participants were asked to name the items on each of the sets of stimuli where they had failed to select the incongruent object. The number of naming errors were found to be very small, but these were nevertheless calculated for each group and contrasted with each groups total number of errors. Three 2×2 chi-squared tests (with Yates Continuity Correction to correct for expected frequencies < 5) failed to reveal a difference between any of the groups ($p = 1.000$). With regards to the justification of their choice of incongruent object, there was only one participant (from the autism group) that omitted to justify a response, so it is unlikely that the clinical groups' difficulty was simply because they were less willing to give a verbal response to have a guess at how the remaining items were related.

Discussion

This attempt to test visuo-conceptual ability led to the inclusion of two conditions. One required the integration of multiple objects (integration condition), and the other required the identification of similarities between objects (category condition). This integration of objects, and detection of similarities, should have left participants with an odd object which they were required to identify. The three groups did not differ in their ability to identify the odd objects in the category condition. Neither did they differ in their response times on this condition. However, the clinical groups were significantly less able to identify some of the odd objects in the integration condition. This difficulty was also reflected in their response times, which were significantly longer for both groups. Moreover, inspection of their data revealed that their longer response times was due to a tendency to take longer on their incorrect items. However, although the clinical groups themselves did not differ in their response times on the integration condition, the autism group were less able to choose the odd object than the Asperger group, although this was significant for just one item.

Unlike the category task, the integration tasks does not afford the opportunity of detecting similarities between objects. Such an arguably bottom-up approach would not facilitate the identification of the type of scene or incongruent object. It is only when the objects are grouped together, a process that is essentially top-down, that the scene and hence incongruent object can be identified. The item analysis supports the notion that central coherence may be playing a role in the integration condition because the clinical groups tended to perform worse on those stimuli which arguably require the greatest degree of integration. For instance, for the stimulus item containing a boy, a ball, a dog, a tin can, and a fence, it arguably requires a greater degree of coherence to interpret the scene as a boy throwing a ball to knock cans off a fence than to interpret it as a boy throwing a ball to or for a dog.

Given that categorization skills have been found to be consistent with mental age in children with autism (Ungerer & Sigman, 1987), it is not surprising that the clinical groups performed near ceiling on of the category stimuli. In addition the ceiling effects were an inevitable consequence of this type of task. Thus, piloting stimuli resulted in a situation where the more difficult categories caused individuals to come up with all sorts of reasons as to why items might be different to the rest (even down to non-category specific information such as size, colour, weight, and height). Since some sets of stimuli could result in more than one item that was different in some way to the others it was necessary to select stimuli where there would not be a great deal of ambiguity as to what was the correct answer (fortunately, the integration condition did not suffer in this way). Although it is quite possible that performance in the category condition might have been limited by a ceiling effect, it is unlikely that this would have appreciably altered our findings. Thus, the clinical groups were significantly impaired on some of the items in the integration condition relative to the normal group and relative to their own performance on the category condition. In addition, the category condition lived up to our main requirement, which was to check whether participants could understand that the core requirement of this experiment was to select the incongruent item. As a check for fatigue, attention, and motivation, the category condition may have been less useful, but even here we would expect to see a drop in performance if any of these factors were playing an appreciable role.

The nature of the two conditions suggest that weak central coherence would affect the clinical groups' performance on the integration condition whilst having a lesser effect on the category condition. Moreover, the normal control group's strong coherence could underly the fact that they performed well on both conditions. It is clear from the results that the clinical groups do not achieve the same degree of coherence as their normal control group. The clinical groups' deficit on the integration condition must strengthen our argument that the task requirements of the integration condition (namely visuo-conceptual integration) may be responsible for their poor performance.

In the general discussion we consider some factors which might be thought to influence the clinical groups' performance on this test. However, at this point what the analysis suggests is that the poor performance in integrating objects coherently characterizes the majority of clinical participants. Given that elsewhere some of these same participants were observed to have an enhanced ability to process locally, then there might be a dissociation between global and local processing ability, so that impaired visuo-conceptual processing goes hand in hand with enhanced local or low-level visual processing. This is explored in the next experiment.

EXPERIMENT 2: THE SCENIC TEST

Introduction

We find bread in kitchens, toothbrushes in bathrooms, and wardrobes in bedrooms. People know a great deal about where objects are likely to be found in the world. The Scenic test seeks to exploit this. Like the previous experiment it assesses visuo-conceptual integration, but differs in that whole scenes are presented. The stimuli are black line drawings of normal everyday scenes, except for one easily nameable item which is not reasonably appropriate for the context. This odd object is naturally represented, for example a butterfly resting on the handle of a spade in a snow scene, and preserves natural pictorial proportions.

There are three parts to this test. The first part is an open-ended "description" task, where participants are required to describe out loud the scenes, in any way they like. This condition seeks to examine whether individuals on the autism spectrum can appreciate a scene, not only for what it portrays, but whether they can notice an inappropriate object. Also of interest is whether the clinical groups can provide coherent descriptions and whether they spontaneously concentrate on local or minor details.

The second part is a "context-sensitive" condition, where participants are required to be sensitive to the context. Whereas the previous condition assesses spontaneous behaviour, this condition demands visuo-conceptual integration, since one is required to conceptually integrate the objects in a picture in order to identify the type of scene, and conceptually process the objects within context in order to identify the context-inappropriate object.

The third part is the "locating of objects in space" condition where participants are required to locate a named (incongruent) object as quickly as they can.

This test provides a quantitative measure of how well individuals can identify a scene. It also provides a quantitative measure of individuals' ability to conceptually integrate objects in a scene, in the first case to spontaneously identify an incongruent object, and in the second case to check the plausibility of individual objects within context. If weak central coherence is common to individuals on the autism spectrum, then the central coherence hypothesis would predict that they would be less proficient at mentally synthesizing the objects in the scene so as to identify the type of scene. Also, they would be less likely to spontaneously synthesize objects in context, which would enable them to notice an incongruent object, and they would be less proficient at synthesizing objects in context when instructed to identify an incongruent object. Moreover, the central coherence hypothesis might predict that they would pay preferential attention to local rather than global aspects of a scene, with the result that their descriptions of a scene are more likely to favour local rather than global aspects. Similarly, any such tendency to process locally rather than globally might suggest an enhanced ability to rapidly locate the named odd objects in the "location" condition. Thus, the prediction was made that individuals with either autism or Asperger syndrome would be less proficient at integrating objects to identify a scene and in making spontaneous and enforced judgements about individual objects in context. A further prediction was made that they would locate the named incongruent object more rapidly on the location task and that their descriptions would be less global and hence more local than normal control individuals.

Participants

These were the same as those who took part in the first experiment.

Materials

The stimuli used in this experiment consisted of black line drawings of scenes. These line drawings were drawn on blank white A4 sheets of paper. There were three sets of scenes: the first was the open-ended description task (description condition) containing 8 scenes; the second was the context-sensitive task (context-sensitive condition) containing 2 trial scenes and 12 test scenes; the third was the location of objects in space task (location condition) containing 8 scenes.

The scenes that were used in this experiment were normal everyday scenes, except that they contained an item that was not reasonably appropriate for the context. The position of the odd item within each scene was varied as far as was possible given the limitations imposed by the types of scene being presented. The pictures were presented to 10 individuals in a pilot study. It was found that the types of scene and the odd item in each scene were relatively easy to name and unambiguous. Examples of the stimuli used in this experiment can be found in Figures 1–4. (The full set of stimuli can be obtained from the first author.)

A stopwatch was used both to control the length of time that scenes in the description task were exposed and to time how long it took participants to find or locate odd items. In order to assist in the recording of response times there was a large foolscap-size blank white card, the purpose of which was to cover the test scenes. This card would be rapidly removed and allowed greater accuracy when timing than would be achieved if one turned over the large test scenes and then started the stopwatch. A tape recorder was used to record the participants' verbal descriptions. This enabled the experimenter to accurately transcribe the material at a later date.

Procedure

The experimenter sat opposite the participant for ease of presenting the stimuli. Each set of scenes were placed in separate piles face downwards on the table. The set of scenes currently being presented were placed face up directly in front of the participant.

The trial and test items. The experimenter began with the open-ended description task (description). Participants were told that they were going to see some scenes and that in this task there were no right or wrong answers. They were instructed to describe out loud the scenes in any way they liked—the sort of description they gave was entirely up to them. However, they were instructed to try to say as much as they could, until the experimenter told them to stop, which would be after 45 s had elapsed from the initial presentation of the scene. All participants were shown the tape recorder and it was explained that it was needed to record what they said. No participant felt too uncomfortable about this part of the test being recorded, and all individuals were able to provide descriptions. The tape recorder was unobtrusively stopped after 45 s even if an individual was mid-sentence. All participants received the scenes in a different random order in order to ensure against order effects. An example is shown in Figure 1.

Upon completion of the description condition, participants were given the context-sensitive condition. Unlike the previous condition there were now two trial scenes, which were presented in fixed order. The experimenter turned to the first trial scene and drew the participants' attention to the type of scene being represented—that of a swimming baths. She then drew their attention to the fact that the scene contained something not formally found in a swimming baths—a dolphin. All participants acknowledged both the type of scene and the fact that a dolphin would not normally be found there. The experimenter than turned to the second of the trial scenes (a fish market/stall) and individuals were encouraged to identify what the odd thing was in this scene. The majority of participants (14/17 individuals with Asperger syndrome, 13/17 with autism and all control participants) recognized the bunch of grapes to be out of place and all individuals recognized that it was a fish market. When participants acknowledged both the type of scene and the odd object for the context the experimenter went on to explain the procedure for the test items.

The participants were instructed that they were going to start the test items and that the procedure was identical, except that this time they were going to be timed to see how long it took them to find what the odd item was. It was

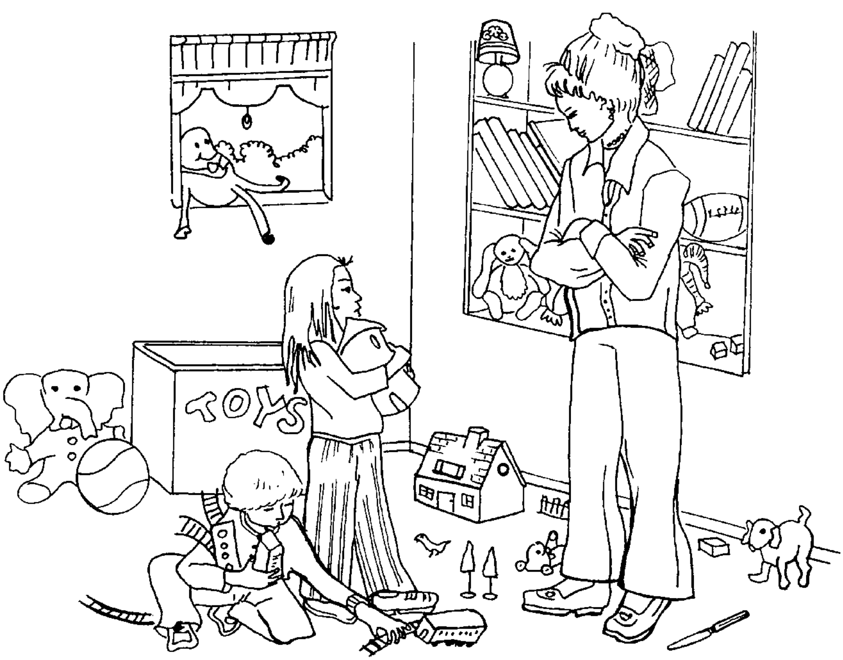

Figure 1. The scenic integration test—description condition example.

mentioned that there was no time limit and that they were to point to or state what the odd item was. They were also told that there was just one odd item in each scene. Furthermore, it was mentioned that they should be familiar with the name of all the odd items that would be used and that they would be able to quite easily name them if they were to see them on their own. They were also told that it would always be a whole item and would never be a part of an item that was odd and it was stressed again at this point that the item must be odd for the context. Also they were told that after they had made their choice they would be asked to say what type of scene it was and again there was no time limit. Participants were told that they would be required to give a one- or two-word statement that stated the type of scene they were viewing. It was emphasized that what the experimenter did not want was things like descriptions, e.g., a "dolphin swimming", for what was really a swimming baths scene, or "people selling", for what was really a fish market scene. They were instructed that whereas they could amend their statement of the "type of scene", they could not amend their choice of odd item, because otherwise their response times would be distorted. All participants fully understood what they had to do, so the experimenter proceeded with the test scenes. The test items were presented in a different random order to each participant, so as to prevent there being order effects. Two examples are shown in Figures 2 and 3.

Participants were then given the third test condition, the location condition. It was pointed out again that one of the things was odd for the scene, and that again they were to go through the scenes pointing to the odd item. In addition, the experimenter said that this time they would be told what the odd item was in advance, so their job would simply be to locate the thing as quickly as possible. It was mentioned that they would be timed (although there was no time limit) in order to see how long it took them to find the stated item, and that they were to point to where it was on the page. They were again instructed that they could not amend their choice as their response times would be distorted. The test scenes were again presented in a different random order to each participant, this again being to prevent order effects. An example is shown in Figure 4.

The location condition was always presented last (after the other two conditions). Being presented last would make it a good control for problems with fatigue, which is presumed to result in slower response times and/or lower accuracy scores in later rather than earlier conditions. Furthermore, this condition would indirectly clarify whether any difficulty portrayed by the clinical groups on the "context-sensitive" task, was due to the clinical groups having difficulty in locating objects in space. The location condition also enabled the experimenter to assess whether the clinical groups were being impulsive. Impulsive responses would not only result in faster response times, particularly on failed items, but would also increase the likelihood of errors.

It was not possible to counterbalance the three conditions. This was because the description condition had to be presented first since the primary reason for

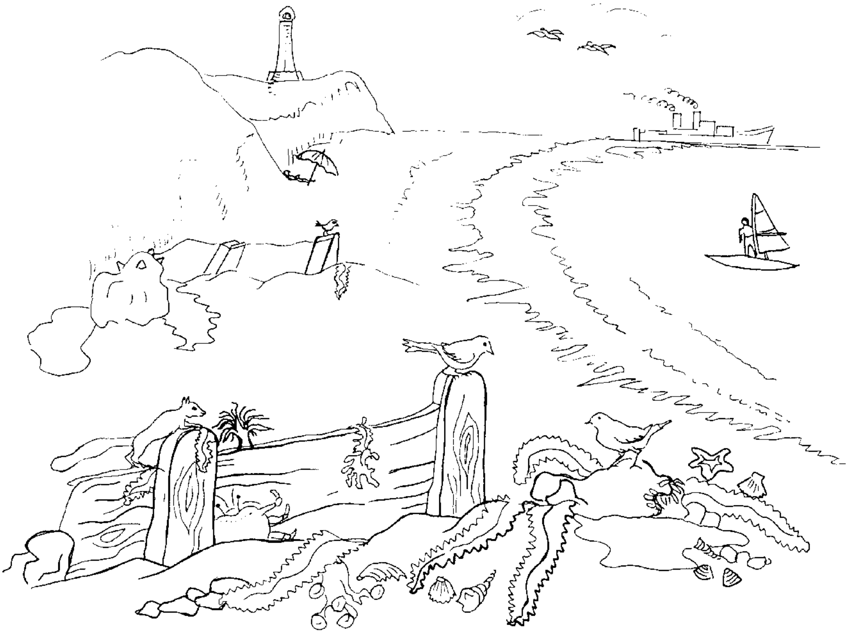

Figure 2. The scenic integration test—context-sensitive condition example.

Figure 3. The scenic integration test—context-sensitive condition example.

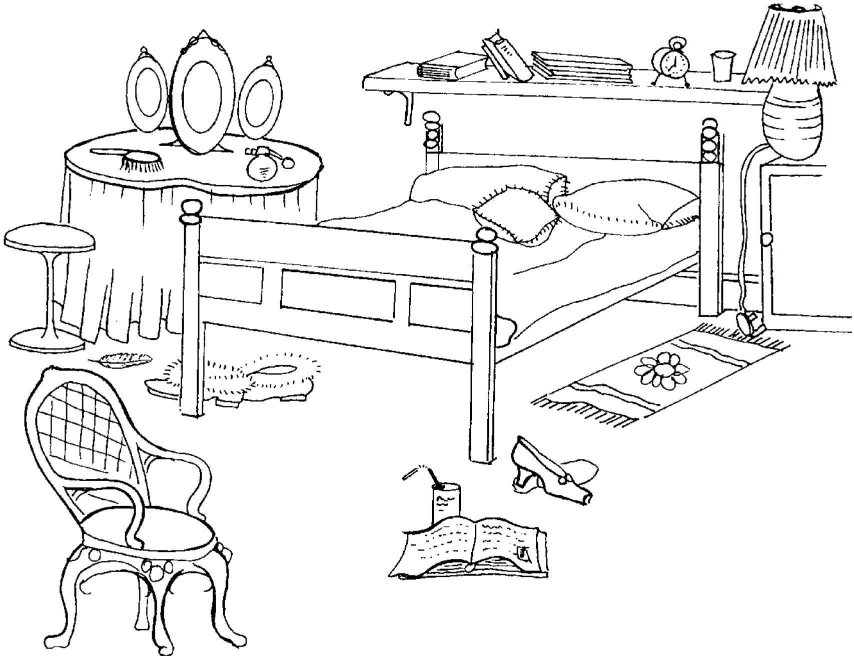

Figure 4. The scenic integration test—location condition example.

conducting this task was to capture the clinical groups' processing preferences. It was felt that there was a strong possibility that the clinical groups would mention the odd item because they are renowned for their tendency to concentrate on local detail. However, for the same reason it was thought that the clinical participants would be less likely to mention its inappropriateness for its context if they tended to focus, more on details and had the predicted problem of being sensitive to context. On the other hand, if control participants had strong central coherence they might be drawn to more global aspects of the scene. Also, given their greater social awareness they might also be drawn to the emotional expression on people's faces. Moreover, with their known sensitivity to context it was predicted that they would be much more likely if they did mention the odd item to mention its inappropriateness.

Object naming. To check if poor performance was due to a lack of familiarity with the odd item portrayed, the experimenter at the end of the test pointed to and got the participant to name the odd item in each of the scenes where they had failed to mention the odd object or had failed to select or locate the odd object.

Scoring on the description task. The first analysis was a quantitative measure on the amount participants had to say, since it was predicted that the clinical groups might have less to say about the scenes than the normal control group. The second analysis made use of a proportionate measure that took account of how much was said by participants. Thus, this analysis aimed to assess how many times each group mentioned the odd object and this was analysed as a proportion of the total number or words uttered by each group. The third analysis was aimed at assessing how many times the odd object's oddity for its context was remarked upon, and again this was analysed as a proportion of the total number of words uttered by each group, and as a proportion of the number of times the odd object was mentioned (either merely mentioned, or mentioned and remarked upon as being odd).

Scoring on the context-sensitive task. There were three types of performance measure: the number of odd objects correctly identified; the number of correct identifications of the scenes; the time the participant took to identify what he/she believed was the odd object.

The scenes were designed to be at roughly three levels of difficulty: very easy to identify scenes such as a kitchen, slightly more difficult to identify scenes such as an autumn scene, and very difficult to identify scenes such as a china and glassware shop (the china and glassware shop is difficult for two reasons: first because the schema is less familiar and hence more difficult to recognize than that of say a kitchen, and, second, because the picture contains a vast number of items and even some unusual glassware items, whilst the inappropriate object is an electric kettle).

Stating the type of scene is to some degree subjective, but given a definite set of instructions to the participants it was possible using specifically defined criteria to rate each individual's statement with a relatively high degree of accuracy. The following criteria (with examples) was used to decide on whether the scene had been stated correctly or incorrectly.

(1) A correct contextual statement would be one that demonstrates a full appreciation of the type of theme being portrayed, e.g., office, play area, desert, and winter or snow scene.

(2) An incorrect contextual statement would be one that fails to demonstrate a full appreciation of the type of theme being portrayed, e.g., those that meet the following criteria would be classified as incorrect:

- a description of the scene (this would go against task instructions), e.g., "raking leaves" for an autumn scene; "people walking" for a street scene
- making an analogy, e.g., "Christmas card scene" for a snow scene
- a global description that fails to demonstrate a full appreciation of the underlying theme being demonstrated, e.g., "garden scene" for an autumn scene and "coastal scene" for a seaside scene.

Stating the type of scene is clearly subjective, so validation of the scoring was necessary. The score sheets were given to a second rater, who was blind to the identity and diagnostic status of the participants and naive to the hypothesis being tested. The degree of agreement across the 12 types of scene was 98%.

The recording of response time was achieved as follows: The stopwatch was started as soon as the test pile's covering card was (rapidly) removed from the top of the pile of stimuli. As soon as the participant pointed to or reported what the odd item was, the experimenter stopped the stopwatch and recorded the choice of object and the response time.

Participants were asked to say what type of scene they thought was being presented. This scene remained in view whilst they made their response and amendments were permitted so as not to penalize participants who might be impulsive, anxious, or had difficulty in communicating their thoughts.

Scoring on the location condition. There were two types of performance measure—the number of odd objects correctly located and the time taken to locate these. As before, the stopwatch was started as soon as the test pile's covering card was removed from the top of the pile of stimuli. As soon as the participant had located the odd item the experimenter stopped the stopwatch and recorded the name of the object located and the response time.

Results

Description task

Amount said. Since it was predicted that the clinical groups would find the description task more difficult than the normal control group, a one-way ANOVA was performed on the number of words uttered by each group. The three groups differed in the number of utterances made, $F(2, 48) = 4.67, p < .05$, and the source of this difference was investigated further using t-tests. Planned contrasts indicated that the autism and Asperger groups were significantly different to that of the normal control group, t aut. $(32) = 3.30, p < .01$; t Asp. $(32) = 2.00, p = .05$. The clinical groups did not themselves differ, $t(32) = 0.77, p = .45$. The mean number of words uttered by the normal, autism, and Asperger groups was 683, 551, and 566 words respectively (SD: normal group 125; autism group 107; Asperger group 114).

Odd object. The number of times the odd object was mentioned was examined for each group as a proportion of each group's total number of words uttered. Three 2×2 chi-squared tests were conducted to see whether the groups differed in the number of times they mentioned the odd object. The analysis revealed that the clinical groups differed significantly from the normal control group, Pearson, χ^2 aut. $(1) = 14.37, p < .001$; χ^2 Asp. $(1) = 5.25, p < .05$. However, the two clinical groups did not themselves differ, χ^2 $(1) = 2.52$, $p = .11$.

The clinical groups mentioned the odd object much less than the normal group (autism group 31; Asperger group 48; normal group 86).

In order to assess the participant' (spontaneous) sensitivity to context, the number of times participants in each group spontaneously mentioned the odd object's oddity for its context was examined. This was done in two ways: first, as a proportion of each group's total number of words uttered, and, second, as a proportion of each group's total number of times the odd object was mentioned. The reason for these two ways of analysing the data being adopted will be explained. It would seem intuitively correct to analyse the remarks as a proportion of the words uttered, since such a measure is tied directly to participants' output and therefore does not penalize the clinical groups for saying less. However, because the clinical groups tended to start their descriptions in the same way as the control group, beginning with describing what the characters were doing, but found this more difficult with the result that they had less time left to talk about local details, it was necessary to analyse their remarks as a proportion of the number of times the odd object was mentioned (either merely mentioned, or mentioned and remarked upon as being odd). This latter measure was not dependent upon the way of describing the scenes or the time limit.

Three 2×2 chi-squared tests were conducted for each way of analysing the data. The analysis as a proportion of the total number of words uttered revealed that the two clinical groups did differ significantly from their normal control group, Pearson, χ^2 aut. $(1) = 28.52$, $p < .001$; χ^2 Asp. $(1) = 22.36$, $p < .001$, although the clinical groups themselves did not differ (Yates Continuity Correction to correct for expected frequencies < 5), $\chi^2\ (1) = 0.71$, $p > .05$. The analysis as a proportion of the total number of times the odd object was mentioned revealed that the two clinical groups again differed significantly from their normal control group, Pearson, χ^2 aut. $(1) = 8.80$, $p < .01$; χ^2 Asp. $(1) = 9.35$, $p < .01$, although again the clinical groups themselves did not differ (Yates Continuity Correction for expected frequencies < 5), $\chi^2\ (1) = 0.05$, $p > .05$. The clinical groups spontaneously remarked upon the oddity of the odd object much less than the normal group (autism group 3 times; Asperger group 7 times; normal group 46 times).

Context-sensitive task

There were 12 scenes. The identification of objects from correctly identified scenes was considered to be the only fair measure of the ability of participants to choose the odd object. This is because identifying an object as inappropriate for a scene requires an appreciation of what the scene actually is in the first place. The object identification score was achieved as follows, e.g., if there were 6 correctly identified odd objects out of 9 correctly identified scenes, then participants would be given a score of 66% on this particular measure. Participants were given a percentage accuracy score for their ability to identify the scene. A percentage score was taken so as to compare the identification of scenes with the identification of objects. The mean percentage scores and mean response time for each of the groups is shown in Table 4.

Accuracy in identifying the scenes. A one-way ANOVA was performed, which revealed, as predicted, a group difference in identifying scenes, $F(2, 48) = 9.69$, $p < .001$. The source of the group difference was investigated further using *t*-tests. Planned contrasts indicated that the autism and Asperger groups were significantly different to that of the normal control group, t aut. $(48) = 4.29$, $p < .001$; t Asp. $(48) = 3.01$, $p < .01$, although the clinical groups did not differ from each other, $t(48) = 1.27$, $p = .21$.

It is important to determine whether the clinical groups' impaired ability to identify scenes is determined by just a few individuals in each group or characterized the majority. Thus, the number of participants in each group scoring above (and below) the control group mean was calculated. This was compared to the number of participants in the normal control group scoring above (and below) their mean. The analysis revealed that the clinical groups differed significantly from the normal control group, χ^2 aut. $(1) = 12.14$, $p < .001$; χ^2 Asp. $(1) = 9.66$, $p < .01$.

Accuracy in identifying the incongruent objects. A one-way ANOVA revealed a group difference in detecting the incongruent object, $F(2, 48) = 11.62$, $p < .001$. The source of the group difference was investigated further using *t*-tests. Planned contrasts indicated that the autism and Asperger groups were significantly different to that of the normal control group, t aut. $(48) =$

TABLE 4
Mean percentage accuracy scores and response time

Participant group	*Accuracy*		*Response time*
(n = 17)	*Odd object*	*Scene*	*Odd object*
Normal			
Mean	86.05	88.24	17.72
SD	11.96	10.23	7.46
Range	(66.7–100)	(66.7–100)	(8–32.6)
Autism			
Mean	59.49	70.10	28.60
SD	20.34	13.20	7.30
Range	(25–87.5)	(50–91.7)	(16.6–43.2)
Asperger			
Mean	69.51	75.49	26.04
SD	15.27	13.33	7.27
Range	(44.4–100)	(50–100)	(12.2–38.2)

4.77, $p < .001$; t Asp. $(48) = 2.97$, $p = .01$. The clinical groups did not themselves differ, but their scores approached significance, $t(48) = 1.80$, $p = .08$.

A further analysis was conducted to assess whether the clinical groups' impairment in identifying the odd object was determined by just a few individuals in each group or was more widespread. To this end the mean percentage of the normal group's ability to correctly identify the odd object (as a proportion of their correctly identified scenes only) was calculated. The number of participants in this group performing above and below this percentage was then calculated. Then the number of clinical participants in each group also performing above and below this percentage for their correctly identified objects (as a proportion of their correctly identified scenes) was also calculated. The clinical groups' number of participants was then compared to that of the normal group. The analysis revealed that the autism group differed significantly from the normal control group (Yates Continuity Correction to correct for expected frequencies < 5), χ^2 (1) = 4.09, $p < .05$, although the Asperger group did not differ from the normal control group (Yates Continuity Correction to correct for expected frequencies < 5), χ^2 (1) = 2.42, $p = .12$.

Response time for selecting the incongruent item. A one-way ANOVA revealed a significant difference between the groups' response times, $F(2, 48) = 9.69$, $p < .001$. Since no predictions had been made about participants' response times, the source of the interaction effect was investigated further using post hoc Newman-Keuls tests. Post hoc comparisons revealed that response times for the clinical groups were significantly different to that of the normal control group ($p < .01$ for both groups) although the clinical groups did not themselves differ ($p = .31$).

Location task

As one would expect, none of the groups had any difficulty in locating the odd item; none of the Asperger individuals made any errors, and the autism and normal groups made only 1 and 2 errors respectively. The errors that did occur were as a result of three individuals making a very rapid response and not being careful (presumably because of the instruction to find a specifically named object as quickly as they could).

The ability to locate objects in space was also assessed in terms of speed of response. A one-way ANOVA failed to reveal a group difference, $F(2, 48) = 1.55$, $p = .22$. However, there was a tendency for the clinical groups to locate the named object more rapidly (in seconds: autism group 1.54, Asperger group 1.70, and normal group 1.81; SD: autism group 0.39, Asperger group 0.53, and normal group 0.44).

Object naming and omissions. None of the participants had any difficulty in naming the odd objects in the description and location conditions. However, there was a small number of naming errors on the context-sensitive condition. The number of naming errors were calculated for each group, and contrasted with each group's total number of scenes where the odd object had not been identified. Three 2 × 2 chi-squared tests (with Yates Continuity Correction to correct for expected frequencies < 5) were conducted. The analysis revealed no significant difference between any of the groups ($p \geq .68$).

No participant omitted to provide an identification of the type of scene portrayed. However, there were a small number of omissions when it came to detecting the inappropriate object. These were examined. Three 2 × 2 chi-squared tests (with Yates Continuity Correction to correct for low expected frequencies) failed to reveal a difference between groups ($p \geq .48$).

Discussion

The Scenic test's three conditions will be discussed, where we begin by discussing the final condition first, since this condition is of relevance to the discussion of the other two conditions.

Location condition. The three groups did not differ in terms of accuracy in locating the named objects. However, contrary to predictions the clinical groups were not significantly faster at locating these objects. This seems to go against findings which report enhanced and rapid disembedding on visuo-perceptual tasks such as the Embedded Figures test (Jolliffe & Baron-Cohen, 1997; Shah & Frith, 1983). It might be that on the latter type of task the difference is due to the normal participant's inability to resist processing in a detrimental holistic manner. However, on our task the clinical groups approached it in the same way as their normal control group; they tended to scan the scene. Although the clinical groups' response times were not significantly faster, they did have a tendency to be faster. This can possibly be explained by the fact that they were not slowed to the same extent as normal individuals, who always tended to look first where they expected the object to be in the picture, rather than where it actually was. For example, they tended to look first at the top of the scene in anticipation of locating the crescent moon in the upper portion of the picture when in fact it was not there.

Description condition. As predicted both groups of clinical participants had significantly less to say about the scenes than their normal control group. Whether the finding reported here reflects a reluctance to say much, a difficulty in generating things to say, a problem in interpreting the scenes, or a combination of these factors is difficult to say, but will not be pursued further because it is of less relevance to the questions we seek to address.

The clinical groups were observed to approach the task like their normal control group—that is, they tended to start by describing the characters, their expressions, and what they were doing, and occasionally they would start by stating the type of scene. However, it seemed that the clinical participants found giving descriptions quite demanding, since there was very little time left for them to start describing more local details, whereas their normal control participants seem to have more time at their disposal in which to do this. Not surprisingly, the clinical groups mentioned the odd object significantly less than their normal control group (this is irrespective of whether its incongruency was commented upon). This was contrary to our predictions and suggests that they were not relying on a purely local strategy. Whether the clinical groups' difficulty was one of describing the characters, their expressions and what these characters were doing is not entirely clear, since they were observed to show other anomalises of interpretation. These will be briefly mentioned.

Whilst only one individual with autism (on only one scene) failed to give some sort of interpretation, what was most striking was the gross misinterpretations of the scenes. Thus, the family clearing away snow after a heavy snowfall was variously described as digging the garden over, digging up vegetables, planting things, and, in the case of one man with autism, as a family building a dyke in order to keep the sea out. He described the water coming in as high as the house and that they were trying to desperately keep it out by building a sand dyke. Although striking, these errors were few in number: 3 for the Asperger group, 7 for the autism group, and no normal individual made such a misjudgement. This type of error was so rare one cannot come to any strong conclusions about these individuals having perceptual deficits, particularly since other experiments have found their object recognition skills to unimpaired (Jolliffe & Baron-Cohen, 1999a; Mottron & Belleville, 1993). However, one can speculate on why such striking errors could have been made. Such errors could have occurred because of a failure to integrate the different objects in a scene. Thus, for the family clearing away snow after a heavy snow fall, participants who interpret it as a family planting, or trying to keep the sea out would be failing to integrate the Christmas wreath on the door of the house and the fact that snow covers much of what is seen in the picture.

This apparent impairment in integrating elements was again observed in their inability to notice the odd objects' inappropriateness for their context. Thus, the clinical groups mentioned the odd objects' inappropriateness significantly less often than the normal control group, and this was both as a proportion of the number of words said and as a proportion of the total number of times the odd object was mentioned. This finding is in line with Frith's (1989) prediction of a global impairment. However, the clinical groups themselves did not differ from each other in the extent to which they inferred the odd objects' oddity.

There were only a few participants exhibiting anomalies in how they used language to describe the scenes so these anomalies were not analysed. However, they will be discussed as they are of relevance to Frith's central coherence hypothesis. A couple of the participants in the Asperger group and several in the autism group provided descriptions that lacked coherence. That is, their description failed to make full sense, so that a listener or reader could not easily derive an orderly flow of meaning. This finding is reminiscent of the less coherent and less complex narratives in family members of children with autism (Landa, Folstein, & Isaacs, 1991). This tendency to provide less coherent narratives by some of the clinical participants was largely because their accounts lacked cohesion, i.e., specific content or function words such as the use of pronouns. Although individuals with Asperger syndrome tended to use more cohesive links than those with autism, they as a consequence made more cohesive errors. The errors made by both clinical groups tended to be a consequence of the poor use of referential devices such as "additioning references", i.e., they sometimes failed to mention people in the scene in an unambiguous way and said things like "the other person" when there were in fact two others, and this "other" had not been previously specified. These observations are in line with the findings that children with autism have been found to make few cohesive links, whereas those with Asperger syndrome make more, but as a consequence they make more errors in the form of additioning references (Baltaxe, Russell, D'Angiola, & Simmons, 1995; Fine, Bartolucci, Szatmari, & Ginsberg, 1994; Tager-Flusberg, 1995).

The most common error that the clinical groups made was adding new information that had no explicit or implicit relation to previous information. For instance, they would make unclear bridging inferences. A clear one would be "There are lots of shoes. They are arranged on the shelf." An unclear one would be "There are some shoes. There is a shelf." They also had a tendency to make ambiguous and sometimes contradictory statements. This is in line with other findings that suggest that children with autism tend to make unclear bridging inferences (Fine et al., 1994) and tend to produce more ambiguous statements (Loveland, McEvoy, Kelley, & Tunali, 1990).

The clinical groups had a tendency not to explain the relationship between the elements in the scene to the same degree as their normal control group, who for example made more of the following types of statement: "the old lady linked her arm into her daughter's arm, so she did not fall and hurt herself". The fact that the clinical groups used fewer explanatory statements may be related to their tendency to say less and use shorter sentences. This absence of explanatory statements has been found in children with autism when asked to narrate a wordless picture book (Tager-Flusberg, 1995).

Context-sensitive condition. The clinical groups were significantly impaired in their ability to identify the types of scene and the odd objects. Their difficulties were reflected not only in their accuracy scores, which were significantly lower for both groups, but in their response times, which were significantly longer for both groups. Although the clinical groups themselves did not differ in any of these respects, there was a non-significant tendency for the autism group to be less able to identify the incongruent object than the Asperger group.

The results seem to suggest that the clinical groups had greater difficulty in conceptually integrating the objects in context. This can be seen as stemming from weak central coherence. On this condition, a difficulty in conceptually integrating objects in context characterizes the majority of those with autism, and to some degree characterizes those with the related condition of Asperger syndrome. Thus, for the Asperger group, an impairment in identifying the scenes characterized the majority of this group, and an impairment in detecting the context-inappropriate objects characterized a large proportion of this group, this proportion approaching significance.

The results of this condition, in conjunction with the results of the other two conditions, hint at a dissociation between global and local processing ability. Thus, in this condition, an impairment in identifying an object's inappropriateness for its context—a global impairment—went hand in hand with a lack of evidence to suggest an enhanced ability to detect a local detail on the location condition. Similarly, in the description condition, an impairment in noticing an object's inappropriateness for its context—a global impairment—went hand in hand with a failure to find evidence of local processing on this same condition. Thus, it seems that impaired visuo-conceptual processing need not go hand in hand with enhanced local (visuo-perceptual) processing and that this finding can be observed whether one uses an open-ended task (like the description task) or a forced choice task (like the context-sensitive and location tasks). Again the apparent dissociation between conceptual (global) processing and perceptual processing might be due to enhanced local processing only being observed on those tasks where normal subjects find it almost impossible to perform in a non-holistic manner (such as the Embedded Figures test and the Block Design test). Our failure to find enhanced local processing might stem from a failure to force the controls to employ a detrimental holistic processing strategy.

GENERAL DISCUSSION

Various factors might be put forward in order to try and account for the clinical groups' lack of proficiency on these tests. Perhaps the most obvious pertains to schematic knowledge. It is well known that schematic knowledge (situation specific knowledge or experience) is brought to bear on information and influences how one perceives information. Therefore, was the problem for the clinical groups one of deficient situation specific knowledge or experience, or was the problem one of applying this knowledge? According to Wing (1996), weak central coherence could impair the development of one's knowledge base. However, it is equally possible (as Wing recognized) that weak central coherence might impair one's ability to see the relevance of one's knowledge and also integrate this. Could it be the case that weak central coherence could both hamper the development of one's knowledge base as well as inhibit the use of knowledge? It is our belief that the clinical groups did have the required knowledge given the simplicity of the scenes presented and given that the errors they made tended to be on stimuli which required the greatest degree of coherence rather than being schematically more demanding. Moreover, it is evident that this coherence or integration problem can be seen on a visual task which does not require any schematic knowledge (Jolliffe & Baron-Cohen, 1999a). In our view it is not a schematic deficit but is more likely to be the coherence nature of the task that is causing the problem.

Another issue is that this task despite being visual can be performed linguistically. Thus, it is possible that some sort of verbal mediation might underly performance. Future research will need to address the issue since verbal mediation is an inherent problem with many visual tasks. However, of interest is the finding that these same participants were impaired on a similar task that did not lend itself to use one's linguistic ability. On this task participants were required to conceptually integrate fragments of an object in order to identify the object (Jolliffe & Baron-Cohen, 1999a).

Another factor underlying these experiments is that participants need to reason as to the degree of appropriateness of the objects or coherence of their scenes. Therefore, failures to check plausibility or revise one's initial choice would affect performance. However, the significantly longer response times on these experiments seems to suggest more rather than less processing, which goes against the idea that participants stop processing after having integrated objects in one particular way.

Another factor which could conceivably effect performance is a problem with mental disengagement (Hughes & Russell, 1993; Russell, Mauthner, Sharpe, & Tidswell, 1991). However, a number of studies demonstrate that children with autism can mentally disengage (see Baron-Cohen, 1989b; Hobson, 1984; Tan & Harris, 1991 on visual perspective taking tasks). Moreover, the experiments presented here demonstrate that these adults with autism or Asperger syndrome have an ability to mentally disengage; in the description condition their descriptions shifted from one aspect of the picture to another, in the location condition they rapidly scanned the whole picture, whereas on the category control task it was essential when looking for similarities to disengage from one item to the next. Another factor which would affect performance is a tendency to fixate on details (Rimland, 1971) and process locally (Frith, 1989; Jolliffe & Baron-Cohen, 1997; Shah & Frith, 1983, 1993). However, the location condition never demonstrated any evidence of a tendency to be drawn to details, whereas the description task never demonstrated any evidence of a tendency to process locally or focus on details.

It is unlikely that the clinical groups' poor performance is a secondary consequence of their being poorly motivated. Several results and observations argue against this hypothesis. First, they were not impaired on the control tasks. Second, despite their less proficient overall performance, they were observed to respond to the varying degrees of difficulty of the stimuli in much the same way as their control group. Third, in the Scenic test the clinical participants' errors suggests motivated rather than unmotivated behaviour, since rather than selecting any item as being odd for the context, when they were having difficulty in finding the incongruent object they would spend a great deal of time trying to select a more unusual but nevertheless context-appropriate object. For instance, in the snow scene (Figure 3), they would select the hanging bird's feeder or the snowman's hat instead of the butterfly. Such evidence leads one to conclude that the clinical groups' difficulty in integrating information is not likely to be due to problems in motivation. The evidence presented previously when taken together also rules out inattention and fatigue playing a role.

Another factor which might be suggested as contributing to their poor performance is that of impulsivity. However, this is unlikely, because the clinical groups (like their control group) tended to respond faster on the stimuli they could identify easily, but like their control group, they tended to respond slower on the stimuli they could not identify easily. Moreover, on both experiments they took significantly longer to make their response than their normal control group, which suggests prolonged thinking times rather than impulsive responding.

The clinical groups' difficulties are unlikely to be due to a memory problem, because in both experiments the stimulus materials remained in view until participants had completed their response. Neither was working memory likely to be a problem, as evidenced by their good performance on the category control task, which required the similarities between items to be held in working memory. What the analysis did suggest is that participants were familiar with the objects employed as demonstrated by their ability to correctly label the objects used in these experiments and that omissions could not account for group differences. The clinical groups unimpaired performance in locating the named (odd) objects in the location tasks suggests that they do not have a problem in locating objects in space.

The ability to integrate objects in context involves various other skills such as visual perception and the ability to discriminate details. Given the clinical groups' relative strength on perceptual tasks such as the Raven's matrices (Raven, Court, & Raven, 1992), it does not seem that visual perception *per se* is a problem for them. Furthermore, it is unlikely given their well-known eye for detail that they overlooked details.

One of the factors that makes this test different from many other tests, and straightforward reasoning tests, is its requirement of understanding what is context-inappropriate, rather than the more straightforward retrieval of factual information, and the usual straightforward questions testing practical reasoning. Thus, this test seems to elicit impairments in the ability to interpret objects within context. A not too dissimilar test is the Absurdities subtest of the Stanford-Binet intelligence tests (Terman & Merrill, 1960, 1973; Thorndike, Hagen, & Sattler, 1986). This presents pictorial (and linguistic) items of increasing difficulty, which depict a silly, illogical, impractical, or improbable situation (e.g., a lady and man sitting out on the porch in the rain). This test, like the scenic integration test presented here, requires not only a knowledge of what is different or illogical, but requires the ability to integrate information so as to make an inference. It has been found that participants with autism perform significantly worse than their control groups on the Absurdities subtest, and this is found to be the case both with normally intelligent adults with autism (Rumsey & Hamburger, 1988) and learning-disabled children with autism (Carpentieri & Morgan, 1994).

The clinical participants found giving a description of a visual scene quite demanding. A couple of the participants in the Asperger group and several in the autism group provided descriptions that lacked coherence. It seems attractive from the nature of errors made and control measures employed to conclude that such errors as did occur, occurred because of a failure to integrate the different objects in a scene. Clearly the clinical groups' impaired ability to describe visual scenes and interpret them is a fruitful line of inquiry for future research. Moreover, it would seem that future work should concentrate on detecting weak central coherence with real scenes rather than with the relatively unnatural stimuli which we used. Future research also needs to explore how these visuo-conceptual deficits relate to the behavioural characteristics of autism and Asperger syndrome, including the islets of ability, idiot savant syndrome, and sensory over- and under-sensitivity.

In the Scenic test's location condition there was no evidence of the clinical groups showing enhanced local processing ability. This failure to find a group difference for local processing might stem from a failure to force normal controls to employ a detrimental holistic processing strategy of the sort we see in the Embedded Figure test and Block Design test. Of further interest is that the normal individuals tended to try and gauge where they thought the object would be in a picture, rather than where it actually was, an approach that was less likely to be observed in the clinical groups. Thus, "meaning" could be argued not to take precedence for them in the same way that it did for the normal control group. Future research might address this issue directly, perhaps by contrasting the time it takes to locate an object as a function of whether the object is in its expected place or not, whether the scene is congruent or not, and whether participants are shown the object to be located or just informed on what to look for.

Although both clinical groups demonstrated an impairment at the visuo-conceptual level, there were wide differences in ability both within these groups and in general between the three groups. The fact that the groups were matched on IQ suggests that their difficulties do not stem from general intelligence. This suggests that it might be central coherence ability that is responsible for the variability in performance in the clinical and normal groups (although this variability was not as great in the normal group). However, inspection of the data suggests that the clinical groups were not at the low end of the normal continuum on these tests of visuo-conceptual integration, since only a minority of their scores overlapped with control individuals. The clinical participants seemed to possess a real impairment. However, they were not completely insensitive to contextual information, they were just significantly less proficient at conceptualizing it. This relative inefficiency characterized the majority of individuals with autism and to some extent those with Asperger syndrome. Moreover, those with autism performed on the tasks at a consistently poorer level than those with Asperger syndrome. This seems reminiscent of their greater disability in childhood (and sometimes in adulthood) than those with Asperger syndrome.

The central coherence theory is one of several psychological theories of autism. This theory along with theory of mind (TOM) and executive function (EF) theories have attempted to address the characteristics of autism. Each of these theories makes specific predictions and provides particularly good explanations for the different characteristics of autism. However, in our view no single theory stands out as being superior to the others. Moreover, they appear to be conceptually and psychologically related. Thus the term "central coherence" not only taps higher cognitive abilities but is sufficiently broad enough to shade into what is commonly considered to be the executive and TOM domain[4].

It is likely that weak central coherence could exert some impact on both the development of a TOM and in using one's TOM. For instance, weak central coherence might give rise to a critical delay in acquiring mentalizing ability, with the result that the individual may not be able to develop a normal TOM (Happé, 1994b). On the other hand, it may be that weak central coherence limits the usefulness of any mentalizing capacity through an inability to integrate the many different sources of information (e.g., tone of voice, facial expression) that provide clues to the mental states and intentions of others. In support of this, when Happé (1994a) presented TOM questions embedded in a more natural social context, which involved integrating an utterance with its context and thus extracting meaning from the context, she found that even subjects who

[4]The central coherence theory is not alone in its breadth; the term "executive functions" is also broad and shades into not only central coherence but TOM skills (see Jolliffe, 1997 for a discussion).

passed second-order belief tasks showed very marked impairments of mental state attribution.

It is also likely that weak central coherence might play a part in some apparent executive deficits, i.e., some of what has been thought to be an executive problem may in fact be due to weak central coherence. Thus, Happé (1994b) noted that a prepotent response (previously rewarded response or responses with a strong external trigger) can become incorrect as a result of a change of context. Thus, if the context is disregarded, the stimulus would in all likelihood invoke the same (but now incorrect) response, and the individual would appear to be failing to inhibit their actions. Similarly, Ozonoff and Miller (1996) have found that high-functioning adults with autism have difficulty in using context to interpret an indirect request (questions beginning "Can you...") with the result that they tend to fail to inhibit the use of a familiar or overlearned response. It is possible, as Wing (1996) suggested, that a difficulty integrating information inhibits the making of plans and predicting what will happen in the future. Moreover, weak central coherence may play a part in failing to abstract critical information, such as the critical variable on the ID/ED set-shifting task (Turner, 1995).

Although central coherence, theory of mind and executive function theories make predictions about each other, they are also distinct, since they make specific predictions and explain these predictions differently. Thus, each has a unique contribution to make but is not without its problems. For instance, the executive function account appears to explain many of the features of autism, but its universality and specificity is at present questionable. The TOM account neatly explains the social and some of the communicative characteristics of the disorder, but it has little success in explaining the restricted and repetitive behaviours seen in autism. The central coherence account can explain the social, communicative, and repetitive behaviour impairments in autism, along with the islets of ability. However, it seems to do so in a way that is intuitively less direct than the TOM account of some of the social impairments, and the executive function account of some of the repetitive behaviours. Moreover, the central coherence account seems to be a particular type of processing style, which may be neither universal nor specific to autism. Furthermore, although the central coherence explanation of autism provides us with a different and fresh perspective on looking at the strengths and weaknesses of children and adults with autism, it suffers from a degree of over-extension. Thus, it is possible to explain the failure on almost any task or failure in any *real* situation as being the result of a problem in integrating information. This is a major drawback of the theory. However, despite this problem, future research can use a systematic approach to investigate and refine this theory. This will lead to a better understanding of what can and should be explained by the central coherence hypothesis. The fact that participants with autism appear to have weak central coherence does not in any way conflict with the TOM and executive function accounts.

In conclusion, this paper has been concerned with investigating central coherence ability. This was explored with two different visual experiments that sought to assess the integration of objects at the visuo-conceptual level. Both the quantitative and qualitative aspects of responses provide support for Frith's (1989) central coherence hypothesis, which suggests that individuals with autism (and to a lesser extent those with Asperger syndrom) have a diminished ability to integrate visual information in context. Thus, they were not able to identify odd objects when forced to make a choice, something that suggests a competence problem. However, in the description condition of the Scenic test, the clinical groups did not spontaneously comment upon the oddity of the incongruent object, a finding that appears to reflect a non-conscious processing preference. Taken together, the results from the two experiments suggest that the disproportionately greater effort that such individuals need to make in order to process visual material in context results in them tending not to process in context unless instructed to do so or unless they consciously decide to do so.

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders [DSM-III-R]* (3rd Rev. ed.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders [DSM-IV]* (4th ed.). Washington, DC: Author.

Baltaxe, C.A.M., Russell, A., D'Angiola, N., & Simmons, J.Q. (1995). Discourse cohesion in the verbal interactions of individuals with autistic disorder or schizotypal personality disorder. *Australia and New Zealand Journal of Developmental Disabilities*, *20*, 79–96.

Baron-Cohen, S. (1989a). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, *30*, 285–297.

Baron-Cohen, S. (1989b). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology*, *7*, 113–127.

Carpentieri, S.C., & Morgan, S.B. (1994). Brief report: A comparison of patterns of cognitive functioning of autistic and non-autistic retarded children on the Stanford-Binet-fourth edition. *Journal of Autism and Developmental Disorders*, *24*, 215–223.

Dunn, L., Dunn, L., & Whetton, C. (1982). *British picture vocabulary scale*. Windsor, UK: NFER-Nelson.

Fine, J., Bartolucci, G., Szatmari, P., & Ginsberg, G. (1994). Cohesive discourse in pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 315–329.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford, UK: Blackwell.

Happé, F.G.E. (1994a). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 129–154.

Happé, F.G.E. (1994b). *Autism: An introduction to psychological theory*. London: UCL Press.

Happé, F.G.E. (1996). Studying weak central coherence at low levels: Children with autism do not succumb to visual illusions—a research note. *Journal of Child Psychology and Psychiatry*, 37, 873–877.

Happé, F.G.E. (1997). Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology*, 15, 1–12.

Hobson, R.P. (1984). Early childhood autism and the question of egocentrism. *Journal of Autism and Developmental Disorders*, *14*, 85–104.

Hooper, H.E. (1983). *The Hooper visual organisation test manual*. Los Angeles: Western Psychological Services.

Hughes, C., & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: Its implications for theories of autism. *Developmental Psychology*, *29*, 498–510.

Jolliffe, T. (1997). *Central coherence dysfunction in autistic spectrum disorder*. Unpublished PhD thesis, University of Cambridge, UK.

Jolliffe, T., & Baron-Cohen, S. (1997). Are people with autism and Asperger syndrome faster than normal on the embedded figures test? *Journal of Child Psychology and Psychiatry*, *38*, 527–534.

Jolliffe, T., & Baron-Cohen, S. (1999a). *Can adults with high-functioning autism or Asperger syndrome integrate fragments of an object?* Unpublished manuscript, University of Cambridge, UK.

Jolliffe, T., & Baron-Cohen, S. (1999b). *Linguistic processing in high-functioning adults with autism or Asperger syndrome: Can global coherence by achieved?* Unpublished manuscript, University of Cambridge, UK.

Jolliffe, T., & Baron-Cohen, S. (1999c). Linguistic processing in high-functioning adults with autism or Asperger syndrome: Is local coherence impaired? *Cognition*, *71*, 149–185.

Klin, A., Volkmar, F.R., Sparrow, S.S., Cicchetti, D.V., & Rourke, B.P. (1995). Validity and neuropsychological characterisation of Asperger syndrome: Convergence with nonverbal learning disabilities syndrome. *Journal of Child Psychology and Psychiatry*, *36*, 1127–1140.

Landa, R., Folstein, S.E., & Isaacs, C. (1991). Spontaneous narrative-discourse performance of parents of autistic individuals. *Journal of Speech and Hearing Research*, *34*, 1339–1345.

Loveland, K.A., McEvoy, R.E., Kelley, M.L., & Tunali, B. (1990). Narrative story telling in autism and Down's syndrome. *British Journal of Developmental Psychology*, *8*, 9–23.

Maccoby, E.E., & Jacklin, C.N. (1975). *The psychology of sex differences*. London: Oxford University Press.

Mottron, L., & Belleville, S. (1993). Study of perceptual analysis in a high level autistic subject with exceptional graphic abilities. *Brain and Cognition*, *23*, 279–309.

Mottron, L., Belleville, S., & Ménard, E. (1996). *Piecemeal graphic construction in high-functioning autism advocates for perceptual abnormalities*. Poster presented at the 14th biennial ISSBD conference, Quebec City, Canada.

Ozonoff, S., & Miller, J.N. (1996). An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and Language*, *52*, 411–434.

Perner, J., Frith, U., Leslie, A.M., & Leekam, S.R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief and communication. *Child Development*, *60*, 689–700.

Raven, J., Court, J., & Raven, J. (1992). *The standard progressive matrices*. Oxford, UK: Oxford Psychologists Press.

Rimland, B. (1971). The differentiation of childhood pyschoses: An analysis of checklists for 2,218 psychotic children. *Journal of Autism and Childhood Schizophrenia*, *1*, 161–174.

Rumsey, J., & Hamburger, S. (1988). Neuropsychological findings in high-functioning men with infantile autism, residual state. *Journal of Clinical and Experimental Neuropsychology*, *10*, 201–221.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows" task as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, *9*, 331–349.

Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry*, *24*, 613–620.

Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the block design task? *Journal of Child Psychology and Psychiatry*, *34*, 1351–1363.
Snodgrass, J.G., & Vanderwart, M. (1980). A standardised set of 250 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.
Tager-Flusberg, H. (1995). "Once upon a ribbit": Stories narrated by autistic children. *British Journal of Developmental Psychology*, *13*, 45–59.
Tan, J., & Harris, P. (1991). Autistic children understand seeing and wanting. *Development and Psychopathology*, *3*, 163–174.
Terman, L.M., & Merrill, M.A. (1960). *Stanford-Binet intelligence scale form L–M*. Boston: Houghton Mifflin.
Terman, L.M., & Merrill, M.A. (1973). *Stanford-Binet intelligence scale form L–M (3rd ed.)*. Boston: Houghton Mifflin.
Thorndike, R.P., Hagen, E.L., & Sattler, J.M. (1986). *The Stanford-Binet intelligence scale (4th ed.)*. Chicago: Riverside Publishing Company.
Turner, M. (1995). *Repetitive behaviour and cognitive functioning in autism*. Unpublished PhD thesis, University of Cambridge, UK.
Ungerer, J.A., & Sigman, M. (1987). Categorisation of skills and receptive language development in autistic children. *Journal of Autism and Developmental Disorders*, *17*, 3–16.
Voyer, D., Voyer, S., & Bryden, M. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.
Wechsler, D. (1949). *Wechsler intelligence scale for children*. New York: The Psychological Corporation.
Wechsler, D. (1955). *Wechsler adult intelligence scale*. New York: The Psychological Corporation.
Wechsler, D. (1974). *Wechsler intelligence scale for children—revised*. New York: The Psychological Corporation.
Wechsler, D. (1981). *Wechsler adult intelligence scale—revised*. New York: The Psychological Corporation.
Wing, L. (1981). Asperger syndrome: A clinical account. *Psychological Medicine*, *11*, 115–129.
Wing, L. (1996). *The autistic spectrum: A guide for parents and professionals*. London: Constable & Co.
Witkin, H., Oltman, P., Raskin, E., & Karp, S. (1971). *A manual for the embedded figures test*. Palo Alto, CA: Consulting Psychologists Press.
World Health Organization. (1992). *International classification of diseases and related health problems* [ICD-10] (10th ed.). Geneva, Switzerland: Author.

Manuscript received July 1998
Revised manuscript received December 1999