COGNITION

Cognition 71 (1999) 149–185

A test of central coherence theory: linguistic processing in high-functioning adults with autism or Asperger syndrome: is local coherence impaired?

Therese Jolliffe*, Simon Baron-Cohen

Departments of Experimental Psychology and Psychiatry, University of Cambridge, Downing Street, Cambridge, CB2 3EB, UK

Received 15 June 1998; accepted 26 May 1999

Abstract

Central coherence theory (Frith, U., 1989. Autism: Explaining the Enigma. Blackwell, Oxford.) is addressed by exploring linguistic processing in normally intelligent adults with either autism or Asperger syndrome, to test whether local coherence is impaired. Local coherence is the ability to make contextually meaningful connections between linguistic information in short-term or working memory. Experiment 1 demonstrated that individuals with an autism spectrum condition were less likely to use the sentence context spontaneously to provide the context-appropriate pronunciation of a homograph. Experiment 2 presented scenarios which had a situation and outcome which only cohered if a bridging inference was drawn. The clinical groups were less likely to select the most coherent (bridging) inference from competing alternatives. Experiment 3 demonstrated that individuals with an autism spectrum condition were less able to use context to interpret an auditorily presented ambiguous sentence. The findings from Experiments 2 and 3 suggest that individuals with autism or Asperger syndrome have a difficulty in achieving local coherence, while the evidence from Experiment 1 suggests a preference not to strive for coherence. Taken together, these results suggest that individuals with an autism spectrum condition are impaired in achieving local coherence, and they have a preference not to strive for coherence unless instructed to do so, or unless they make a conscious decision to do so. Moreover, the three experiments correlate with one another, which suggests that central coherence may be a unitary force in these different tasks. Of the two clinical groups, the autism group had the greater difficulty in achieving coherence. Possible explanations for the clinical groups' difficulty are explored.

Keywords: Central coherence theory; Asperger syndrome; Autism

* Corresponding author. Fax: +44-1223-333564; *E-mail address*: sb205@cus.cam.ac.uk (T. Jolliffe)

0010-0277/99/$ - see front matter
PII: S0010-0277(99)00022-0

1. Introduction

In autism, phonology, semantics and syntax appear to be intact, i.e. to be reading and mental age appropriate (Bartolucci et al., 1976; Frith and Snowling, 1983; Tager-Flusberg et al., 1990; Minshew et al., 1995). However, despite this, these individuals have difficulty in appreciating 'meaning'. Research into the linguistic impairment in autism has therefore begun to study comprehension of meaning at the level of sentences and beyond. Prior and Hall (1979) demonstrated that individuals with autism are impaired in their comprehension of phrases, but not in their comprehension of single words. Their reading comprehension scores also tend to be lower than their reading accuracy scores (Lockyer and Rutter, 1969; Rutter and Bartak, 1973; Frith and Snowling, 1983). Difficulty in appreciating meaning has been found with spoken language (Tager-Flusberg, 1981a,b), and Frith and Snowling (1983) demonstrated that children with autism are less able to select the story-appropriate word to fill in gaps in a story. Similarly, people with autism are less likely to use sentence context spontaneously to inform the pronunciation of homographs (Frith and Snowling, 1983; Happé, 1997).

The difficulty in using context has been described by Frith (1989) as stemming from a failure of a central system whose job it is to integrate sources of information to establish meaning. She called this 'central coherence' and suggested this is weak in individuals with autism, with the result that they attend only to small bits rather than to large, globally coherent patterns of information. Some evidence for Frith's theory comes from a personal account of a high-functioning individual with autism, who wrote about having difficulty with similar sounding words (e.g. 'bull' and 'ball') (Jolliffe et al., 1992). This suggests a failure to process words in context, since normally ambiguities such as these are resolved by reference to the higher-level context in which the words are embedded. The difficulty experienced by this individual with autism is at least consistent with weak central coherence theory. In the non-verbal domain, weak central coherence is held to account for the superior performance individuals with autism show on the Embedded Figures test (Shah and Frith, 1983; Jolliffe and Baron-Cohen, 1997), and on the Block Design test (Shah and Frith, 1993). Thus, there is growing evidence consistent with central coherence theory.

The three experiments reported in this paper investigate the ability to process language for meaning. In the Homograph test (Experiment 1), participants are tested to see whether they spontaneously give the context-appropriate pronunciation of a homograph. This is an attempt to replicate Happé (1997). In the Local Coherence Inferences test (Experiment 2), participants are required to select the intervening statement which best fits to make two other statements coherent. In the Ambiguous Sentence test (Experiment 3), participants are required to integrate an ambiguous sentence with its context, in order to select the context-appropriate interpretation of the ambiguous sentence. These three tests investigate what psycholinguists call processing of local meaning. 'Local', therefore, according to linguists, is information which is in short-term or working memory at the same time (usually between one to three sentences), while 'global' pertains to larger units of information which cannot be conceived as being in short-term or working memory at one time (usually five or more sentences; see McKoon and Ratcliff, 1992; O'Brien, 1995). This paper investigates whether local coherence is impaired in people with autism or Asperger syndrome. The findings from the three experiments to be presented should be similar if achieving coherence involves central thought processes (Frith, 1989). Thus, if individuals with an autism spectrum condition have weak central coherence, then there should be a pervasive difficulty with all three tasks, irrespective of whether the material is self-read or is presented auditorily.

2. Experiment 1: the Homograph test

A fundamental requirement in comprehending a sentence is that words must be interpreted according to the context of the sentence. This interpretation process is relatively automatic in normal individuals, as demonstrated by the ease with which we recognise the sense of the many ambiguous words used in everyday speech (son-sun, meet-meat, sew-so, pear-pair). Another example is how we tend to disambiguate homographs when we read: words that are spelled the same, but have different pronunciations and meanings. The meaning and pronunciation of homographs depends exclusively on the context in which they are used. Examples include the meaning and pronunciation of words such as tear, bow, lead, row, read, etc.

If there is a problem in processing meaning in people with autism, they should show reading difficulties when it is necessary to take meaningful context into account, but not when reading single words. Frith and Snowling (1983) devised a task presenting homographs in different contexts. Comparing normal children with children with autism or dyslexia who were of the same reading age, they found children with autism to be less likely to process words in context so as to obtain meaning. Rather, children with autism tended to provide the more common interpretation of the homograph, effectively disregarding the context. The finding is all the more surprising when one considers the fact that the homograph appeared after (and never before) the sentence context.

Recently, Happé (1997) used the modified homograph test developed by Snowling and Frith (1986). Using 20 sentences, five homograph words were presented in four conditions: rare or common pronunciation, and before or after sentence context. She looked for an effect of the position of the homograph in the sentence. The rationale was that if individuals use context to determine pronunciation, then homographs appearing later in the sentence (after the context) should be easier to pronounce appropriately compared to those appearing early in the sentence, before disambiguating context is given. Like Frith and Snowling (1983), Happé did not give participants explicit training, so her test tapped processing preferences. She found that people with autism of normal or near-normal intelligence showed a relative failure to process information for context-dependent meaning, and that this was independent of their theory of mind ability.

Experiment 1 reported below sets out to replicate the findings of Happé (1997) and Frith and Snowling (1983), and extend them in two ways. First, both clinical groups were high-functioning, having an average or above average IQ (full scale, verbal and performance). Second, two groups of adults with an autism spectrum condition were tested, participants being distinguished on the basis of whether they had received a diagnosis of autism or Asperger syndrome. This was to test whether any weakness in processing linguistic information in context was a function of early language development, since autism, unlike Asperger syndrome, is characterised by a clinically significant language delay.

If individuals with an autism spectrum condition have weak central coherence, they should show evidence of this in their failure to use the context to provide the context-appropriate pronunciation of a homograph. It was predicted that the individuals with autism would show a failure to integrate the homograph with the sentence context. No predictions were made as to how the group of adults with Asperger syndrome would perform.

2.1. 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[1] 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 autistic symptoms 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 Organisation, 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 head injury. The normal control participants did not have a family history of autism or Asperger syndrome and 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.

[1]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.

Table 1
Participant characteristics. Data are mean ± SD with ranges in parentheses.

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

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.[2]

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 $F(2,48) = 0.59$, $P = 0.56$; VIQ $F(2,48) = 0.51$, $P = 0.60$; PIQ $F(2,48) = 0.58$, $P = 0.57$; and FSIQ $F(2,48) = 0.10$, $P = 0.91$. The sex ratio (male:female) in all three groups was 15:2, reflecting the sex ratio found in these clinical groups in other studies (Wing, 1981; Klin et al., 1995). The groups were closely matched on handedness, with 15 right-handed and two left-handed individuals in the normal and high-functioning autism group, and 14 right-handed and three left-handed in the Asperger group. All participants were born in England and English was their first language. Several individuals within each group were either studying for, or held, formal qualifications such as a university degree or diploma.

2.2. Materials

The stimuli consisted of 20 test sentences and 13 pre-test single words. Each of the 20 test sentences appeared on a separate card. All the 13 pre-test words appeared together on a single card and five of these 13 words were actually the homographs used in the test. Each of the test sentences and the pre-test words were typed on white card. There were four types of sentences: homographs with a rare pronunciation, the target word coming either before or after the sentence context; and homographs with a common pronunciation, again the target word coming either before or after the sentence context. For homographs with a rare pronunciation, these were labelled on the reverse of the card RB for those whose homograph appeared before the sentence

[2] Participants were given first- and second-order theory of mind tests. The first-order task was a version of Perner's (Perner et al., 1989) Smarties task. The second-order task was the ice cream van test of Baron-Cohen (1989). Whereas all participants passed the first-order task, five 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 re-tested on a new variation of the second-order belief task and all were found to pass.

context, and RA for those whose homograph appeared after the sentence context. For homographs with a common pronunciation, these were labelled on the reverse of the card CB for those whose homograph appeared before the sentence context, and CA for those whose homograph appeared after the sentence context. These labels were to facilitate scoring.

The sentences used in this experiment were taken from Happé (1997). None of these sentences referred to mental states or were very social in nature. It was found that the pre-test words were easy to read, and they were presented to check the participant's ability to read the target words. Examples of the test sentences and the 13 pre-test words can be found in the Appendix.

2.3. Procedure

Each participant was tested by themselves in a room free from distractions. The experimenter sat opposite the individual, so she could conceal the names of the homographs which were to be checked off on a score sheet.

2.3.1. The test and pre-test items

The cards were placed face down on the table directly in front of the experimenter. The participant was told that he/she was going to see some words which they were to read aloud. The experimenter handed the individual the first card which was the pre-test list of words. All participants completed reading these words without any difficulty. The experimenter then shuffled the 20 test sentences thoroughly. The sentences were presented in a different random order to ensure that there would be no order effects. Participants were given one card at a time and were asked to read the sentence aloud.

For consistency, the pre-test words were always presented first, followed by the test sentences. This was to ensure that participants were able to pronounce the homographs.

As in the study of Happé (1997), participants were not alerted to the special status of the homographs. It was assumed from the pre-test reading of the words and the fact that all individuals were of normal VIQ, that they were familiar with both uses of the target words. This assumption seems reasonable in light of the fact that there were no reading errors in any of the sentences except on the homographs themselves.

2.3.2. Scoring

Participants were scored on their correct pronunciations, which included corrected attempts. Participants received a score out of five for each of the four conditions, and where a participant corrected their wrong pronunciation this was noted so as to provide data on participants' self-corrections.

Correct pronunciation of the homographs was used as the measure of context sensitivity. This measure constituted an accuracy score, which ranged from 0 to 5 for each of the four conditions. The mean accuracy scores for the three groups, by type of condition, are illustrated in Table 2.

Table 2
Number of homographs pronounced context appropriately. Data are mean ± SD with ranges in parentheses.

Participant group (n = 17)	Common pronunciations		Rare pronunciations	
	Before	After	Before	After
Normal	4.88 ± 0.33 (4–5)	5.00 ± 0.00 (5–5)	4.88 ± 0.33 (4–5)	5.00 ± 0.00 (5–5)
Autism	4.77 ± 0.44 (2–5)	4.88 ± 0.33 (4–5)	3.82 ± 1.19 (2–5)	4.06 ± 0.97 (2–5)
Asperger	4.77 ± 0.44 (4–5)	5.00 ± 0.00 (5–5)	4.29 ± 0.99 (2–5)	4.59 ± 0.71 (3–5)

Due to the unequal and zero variance, non-parametric tests were employed in order to see whether the groups differed on each of the four conditions. Four Kruskal–Wallis tests were performed on the mean accuracy scores for each condition. This analysis revealed a non-significant result on the common pronunciations (CA: $\chi^2 = 0.46$, $P = 0.79$; CB: $\chi^2 = 0.\ 46$, $P = 0.79$). But as predicted, both the rare pronunciations were significant (RA: $\chi^2 = 8.90$, $P = 0.01$; RB: $\chi^2 = 7.06$, $P = 0.03$). The source of the rare conditions' significance was investigated further using Mann–Whitney tests. To reduce the family-wise error rate, the two rare conditions (RA and RB) were combined, as keeping these conditions separate is not particularly meaningful since the clinical groups perform poorly irrespective of the position of the homograph. The rare mean for the autism and Asperger groups was significantly different to that of the normal control group (autism $U = 58$, $P = 0.002$; Asperger $U = 87.5$, $P = 0.049$) although the clinical groups themselves did not differ ($U = 102.5$, $P = 0.150$). Observation of the means (see Table 2) shows the clinical groups to be significantly worse at pronouncing the homograph correctly for its context.

2.4. Discussion

Both groups of clinical participants performed significantly less well on providing the context-appropriate pronunciation of a homograph when the context required a more rare interpretation, e.g. ‘There was a big *tear* in her dress’. However, they performed normally on providing the context-appropriate pronunciation of a homograph when the context required a more common interpretation, e.g. ‘There was a big *tear* in her eye’. The crucial difference between the rare and common interpretations is that a failure to take the context into account would lead to poor performance on rare interpretations, but have no effect on common interpretations.

This inefficiency in integrating a homograph with the sentence context replicates the findings of Frith and Snowling (1983) and Happé (1997), and is consistent with a central coherence theory. However, the finding is at odds with one of Happé’s main findings, which found the position of the homograph in the sentence context to have an effect. Thus, in Happé’s study the normal control children were found to be able to provide the context-appropriate homograph when the context preceded the homograph, but their performance was less reliable when the context followed the homograph. In the study reported here, the normal control group performed well whatever the position of the homograph. Thus, one would assume that their reading ability and experience enabled them to read ahead and thus perform similarly irrespective of the position of the homograph (see Table 2). It may be that in the Happé study, the young normal children lacked the reading experience necessary to be able to use the information following the homograph in order to disambiguate it. However, consistent with Happé's study is the finding that the clinical groups' poorer performance on the rare condition was independent of the position of the homograph. This would seem to suggest a more general weakness in processing information in context. The significant result on the rare condition is not likely to be due to poor self-monitoring, i.e. that the clinical groups perform poorly simply because they fail to monitor their reading so do not correct their wrong pronunciations. This is because even when one uses a strict score which does not allow for self-corrections, we still get a significant result (RA: $\chi^2 = 7.42$, $P = 0.02$; RB: $\chi^2 = 11.27$, $P = 0.00$).

The open-ended nature of the task enabled the participants to choose to process the stimuli, either 'in context' or 'out of context'. The tendency that the clinical groups had of processing information out of context is likely to reflect a processing preference since Snowling and Frith (1986) found that younger and less able individuals with autism could be trained to pronounce these homographs context-appropriately. Thus, Experiment 1 is fully consistent with central coherence theory, because it demonstrates that for those with autism or Asperger syndrome there is a preference not to process self-read material fully for meaning. It remains to be seen whether there could also be an impairment in integrating linguistic information. The following two experiments address this issue. These explore performance on two tasks which are much more demanding, and which require participants to integrate information in context.

3. Experiment 2: local coherence inferences

Inference processing has been a central issue in the psychology of language (Singer, 1993). One type of inference that has received extensive consideration is the bridging inference (Haviland and Clark, 1974). Bridging inferences (or coherence inferences as they are also called) serve to establish connections between clauses, i.e. the current clause and preceding discourse, or stated differently, antecedent and outcome events. Experiment 2 makes use of such bridging inferences. Two or three-sentence scenarios are provided. The first one or two sentences provide a premise or situation, e.g. George left his bath water running. The last sentence provides information that only coheres to the first if an inference is drawn, e.g. George cleared up the mess in the bathroom. Although the last sentence is an inference-inducing sentence, e.g. the bath overflowed, participants are provided with the correct inference along with two distractors, so their task was one of selecting the most coherent of the three inferences presented. The correct inference, when combined with the text sentences, enables the sentences to be treated as an integrated whole. This test provides a quantitative measure of how well individuals can achieve linguistic coherence. As such, it differs from the previous experiment, which tapped spontaneous processing preferences. This test is a further test of central coherence theory, because the theory of Frith (1989) suggests that individuals with an autism spectrum condition have difficulty in deriving meaning and understanding the conceptual links that join pieces of information. Thus, her theory would predict that they would be less proficient at selecting coherent inferences.

3.1. Participants

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

3.2. Materials

The stimuli consisted of one trial item, 18 test items and a post-test reading passage with questions. All the items, including the reading passage and its questions, appeared on separate white cards. The trial card had a pair of sentences, a statement and three choices of answer. The 18 test cards also contained a pair of sentences (and in one case three sentences), a statement and three choices of answer for each pair. The card with the reading paragraph contained just a single paragraph of 151 words in length, whilst its memory and comprehension question appeared on a separate card. The trial item was labelled T and the 18 test items were labelled from 1 to 18, all these cards being labelled on their reverse side. A stopwatch was used in order to time how long it took the participant to determine what the coherent inference was, and also to time reading speed.

Although four of the 18 test sentence pairs contained mental state terms, the majority were not social or required an understanding of mental states. It was found that all the sentences were easy to read. A pilot study showed reasonable agreement on which of the three possible inferences was truly coherent for the situation (context) and outcome. Thus, of the 50 pilot participants, at least two-thirds agreed with the experimenter on each individual item, and overall participants had an accuracy of at least 75%. The position of the coherent inference was completely orthogonal, such that it appeared in each of the three possible positions six times. Examples of the pairs of sentences, their questioning statement and their three choices, and the post-test reading paragraph with its questions can be found in Appendix A.

3.3. Procedure

Each participant was again tested alone in a room free from distractions. The set of cards were placed face downwards on the table directly in front of the experimenter.

3.3.1. The trial item

The participant was told that he/she was going to see some pairs of sentences. It was also explained that he/she was going to see a statement, which was a question about the pair of sentences, along with three choices of answer, all of which had to be read out aloud. The experimenter then handed the participant the trial card so that the experimenter could explain the requirements of the task and to check that individuals understood these requirements. Participants were told that the first sentence of each pair always described a situation, whilst the second described its outcome. Each individual was told that there was a sentence which was missing, and which should be in the middle of the pair, and that this sentence would make the situation and outcome both related and coherent. They were then told that the object of this task was to choose the sentence which best fitted what they had read, that is the sentence which would make the pair of sentences coherent. It was emphasised that they had to choose the most coherent of the sentences presented. On this trial item, participants were given as much help as was necessary and it was carefully explained why one answer was correct or coherent, whilst the others were not or were incoherent. All participants reported that they fully understood the task requirements so the experimenter proceeded with the test items.

3.3.2. The test items

The participants were then told 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. They were instructed that there would be no time limit, but they were to point to or say whether the correct sentence was the top, middle or bottom one. This was a forced choice test and participants were instructed to make a guess if they really did not know what the correct sentence was. The 18 test cards were shuffled and therefore presented in a different random order to each individual. This was done to avoid order effects. Each individual was given one card at a time, and was asked to read aloud what was written on the card. Positive comments only were made throughout.

3.3.3. Control measures

Individual responses were timed so as to assess whether the clinical groups had longer response times (reflecting difficulty) or faster response times (possibly reflecting impulsiveness on failed items) and whether they were slower at reading. Reading speed was assessed by timing participants on the post-test reading passage. Although there were no reading errors in any of the sentences, we checked that individuals could comprehend what they had read, so as to guard against the possibility that participants could fail the test items due to an inability to comprehend what they were reading. Therefore, at the end of the test, the experimenter asked each participant to read the timed reading passage, and without advance warning, this was followed by a memory and comprehension question. They were not, of course, permitted to refer back to the passage. This ensured that any difficulty on the task could be separated into comprehension problems, or problems in recalling what they had read, versus real problems in integrating linguistic information in a coherent way. Furthermore, since this passage with its memory and comprehension question came right at the end of the test, it also served as a control for fatigue, motivational and attentional factors.

3.3.4. Scoring

There were two types of measure for each participant: the time taken to identify what the coherent sentence was, and the accuracy score, i.e. the number of coherent sentences identified. A stopwatch was used for recording response time. Timing was started as soon as the card was placed face down on the table in front of the person. As soon as the participant pointed to, or reported whether the coherent sentence was top, middle or bottom, the experimenter stopped the stopwatch and recorded on the score sheet the response time which was placed next to the cards identity number. Response times were scored in seconds and 100th of a second and participants were given as long as they wanted to try to identify what the correct inference was, but they were not permitted to amend their choice as this would distort response times. The choice of sentence or coherent inference was also noted on the score sheet. Participants received a score out of 18, that is the number of correct inferences. For the post-test reading passage, the reading time was taken to the nearest whole s as fractions of a s were not considered to be particularly useful. This was entered on the score sheet along with the participant's response to the memory and comprehension question.

3.4. Results

The mean accuracy score and mean response time for each of the three groups is shown in Table 3.

3.4.1. Accuracy

The data was approximately normally distributed for each group and there were no obvious outliers, so two one-way ANOVAs were performed, one on each performance measure. A one-way ANOVA revealed, as predicted, that the groups differed significantly in accuracy ($F(2,48) = 11.63$, $P < 0.001$). This group effect was investigated further using t-tests. Planned contrasts indicated that the autism and Asperger groups were significantly less accurate than the normal control group (t autism(48) = 4.82, $P < 0.001$; t Asperger(48) = 2.45, $P < 0.05$) and that the autism group was significantly less accurate than the Asperger group ($t(48) = 2.37$, $P < 0.05$).

3.4.2. Response time

A second one-way ANOVA, this time on response time, also revealed that the

Table 3
Accuracy, response time and reading speed. Data are mean ± SD with ranges in parentheses.

Participant group (n = 17)	Accuracy score out of 18	Response time in s	Reading speed
Normal	15.00 ± 1.58 (13-18)	29.28 ± 21.59 (2.8–67.1)	47.24 ± 5.59 (39–60)
Autism	11.53 ± 2.15 (7–15)	52.63 ± 32.58 (10.0–128.1)	50.71 ± 8.89 (34–64)
Asperger	13.24 ± 2.46 (10-17)	32.21 ± 27.00 (3.1–83.8)	50.12 ± 6.03 (41–61)

groups differed significantly ($F(2,48) = 5.52$, $P < 0.01$). Since no firm predictions had been made about response time, the source of the group effect was investigated further using post-hoc Newman–Keuls tests in order to reduce the family-wise error rate. Post-hoc comparisons revealed that the autism group was significantly slower in making their response than the normal control group ($P < 0.05$), although the Asperger group did not differ significantly from the normal control group or the autism group ($P > 0.05$).

3.4.3. Control measures

It is unlikely that the group difference in accuracy could have been due to problems in comprehending the material, since 15 out of 17 participants in the autism group and 16 out of 17 participants in the normal and Asperger groups passed the reading passage's comprehension question. Similarly, the difference in accuracy is unlikely to be due to memory problems, since all 17 in the Asperger group passed the reading passage's memory control question, as did 16 out of the 17 participants in the autism and normal groups. There was no difference between groups in terms of reading speed ($F(2,48) = 1.20$, $P = 0.31$) (see Table 3).

3.5. Discussion

The clinical groups were significantly impaired in their ability to identify the coherent inference. This again is strong support for the central coherence theory. The impairment was not only reflected in their accuracy scores, which were significantly lower for both groups, but also in their response times, which were longer for both clinical groups and significantly longer for the autism group. The clinical groups themselves differed, the autism group being significantly less able to select the coherent inference than the Asperger group, although these groups did not differ in their response times. This apparent impairment in integration is reminiscent of the finding from Experiment 1, where the clinical groups were significantly less proficient at integrating homographs in context. However, the essential difference between this and the previous experiment is that this experiment required participants to integrate information, thus demonstrating in the clinical groups more than just a processing preference, but rather what appears to be a deficit in integrating linguistic information.

Exactly how coherence inferences are made is a matter of much debate (see Kreunz and MacNealy, 1996 for a review) but it is generally accepted that the ideas presented in different sentences are integrated into a coherent whole (Haviland and Clark, 1974; Clark, 1975; Thorndike, 1976). Thus having read ‘George left the bath water running. George cleared up the mess in the bathroom’, one might infer that the bath had overflowed. According to the validation model of bridging inferences, such inferences are only accepted by the understander when validated not only with reference to the cause and outcome, but on the basis of one's knowledge (Singer et al., 1990; Halldorson et al., 1992; Singer, 1995). However, if existing knowledge plays a part in accepting coherence inferences (whether self-generated or provided) then an important aspect of making coherence inferences is the ability to integrate prior general knowledge with the text. A failure to select the coherent inference could therefore be due to inadequate integration and use of general knowledge during reading, or an impaired knowledge base itself. It is possible, as Frith (1989) argued, that weak central coherence could itself impair the development of one's knowledge base, as well as one's ability to use it.

There are basically two types of knowledge which could be brought to bear on such tasks as that reported here: one is script knowledge[3] (which was avoided as far as was possible) and the other is a more general type of world knowledge[4]. Together, these different types of knowledge facilitate coherence. Thus failures in this task may reflect in a minority of instances a lack of world knowledge or a failure to recognise the event sequence or script. Little is known about the underlying world or script knowledge in people with autism or Asperger syndrome. However, it is unlikely that these normally intelligent adults would be that deficient in this knowledge, and attempts were made to reduce the demands for script knowledge, keeping the number of stimuli that relied on this type of knowledge down to just four items.

Two of the script stimuli were also social (mentalistic) and there were two other social (mentalistic) but non-script items. Of the two non-social (non-mentalistic) script items, one coherent inference actually went against what one would expect from script knowledge. This means that less than one-third of the task stimuli required social and/or script knowledge. It is unclear whether deficient social and script knowledge or a problem in using this knowledge could have affected the clinical groups' performance on this task. However, errors that were made by the clinical groups were evident on all types of stimuli, so the difficulty was not confined to social material or script items. Furthermore, there was an observed tendency for clinical participants to take longer and be less accurate not on those items tapping social (mentalising) and/or script knowledge, but on those which placed the greatest demands on integration to establish higher-level meaning.

The results seem to suggest that, like Experiment 1, it is the need to integrate linguistic information in context that accounts, at least in part, for the clinical groups' impaired performance. The finding that the autism group is significantly more impaired than the Asperger group is reminiscent of their comparatively greater failure to integrate the homograph with its context in Experiment 1. However,

[3]Scripts are usually thought of as static knowledge structures that capture our expectations about common, everyday activities, such as eating in a restaurant or visiting a doctor's consulting room (Schank and Abelson, 1977; Bower et al., 1979). A script is a type of event schema, one that includes knowledge about what will happen in a given situation, and often the order in which the individual events will take place (Mandler, 1984). They are thought to be formed through experience with stereotyped event sequences (Schank and Abelson, 1977; Abelson, 1981).

[4]This general world knowledge includes spatial knowledge (Glenberg et al., 1987; Morrow et al., 1989), temporal knowledge (Anderson et al., 1983) and mentalistic knowledge (Gernsbacher et al., 1992). The present test sought to keep mentalising demands to a minimum, however it still required some ability to take into account the more simple mental states of other individuals. Also, readers may activate physical knowledge, which captures how one physical event precipitates a subsequent physical event, and enabling knowledge, which captures how one state or event can fulfil preconditions that allow the occurrence of a second event (Kreunz and MacNealy, 1996).

whereas the finding in the previous experiment may have reflected a processing preference, this experiment supports a competence problem. This suggests that when one looks at higher-level meaning with linguistic tasks, then even at the local level weak central coherence characterises individuals with either autism or Asperger syndrome. The final experiment further explores the idea of a deficit in integrating linguistic information. It aims to see whether the deficit is apparent with auditorily presented stimuli, using stimuli which do not require sophisticated social knowledge or mentalising ability or a knowledge of scripts. A deficit in integration would be expected if the central coherence deficit is part of 'central' cognition, and not specific to one modality.

4. Experiment 3: the Ambiguous Sentence test

In everyday conversations, ambiguous sentences tend to go unnoticed and misinterpretations are rare (Hoppe and Kess, 1986). The context and topic determine the interpretation of even the most ambiguous sentences. For example, assume that your best friend had just had his eyesight tested, and you were told, 'He bought some glasses'. In this case, you would logically interpret the sentence to refer to spectacles. Now assume that your best friend has recently bought his first house and you were told, 'He bought some glasses.' In this case, the sentence logically refers to drinking glasses.

Experiment 3 makes use of two types of ambiguity – lexical and syntactic ambiguity. An example of lexical ambiguity would be: 'He drew a gun', where the verb could mean 'drawing' or 'pulling out'. An example of syntactic ambiguity would be: 'The man was ready to lift' – here the ambiguity occurs because there is more than one logical relationship which can exist among the words since it is unclear whether the man is ready to lift something, or the man is ready to be lifted.

The Ambiguous Sentences test used here is a novel test. Unlike the previous two experiments, it requires participants to listen to linguistic material. This test consists of pairs of sentences which are presented auditorily. The last sentence of each pair is ambiguous, but can be disambiguated by the first sentence which forms a biasing context[5]. Each ambiguous sentence is presented twice in two differing contexts. One context biases towards a rare interpretation and the other to a common interpretation. Participants are required to treat each sentence pair as a whole so as to answer a question about the ambiguous sentence. The question elicits whether the individual

[5]The context was always presented first, before the ambiguous sentence, because it has been found with high-functioning adults with autism that when ambiguous (misleading) information is presented first and the context follows, participants have difficulty in revising their initial interpretation when required to do so (Ozonoff and Miller, 1996). If the ambiguous sentence was presented first and its context second, then there is a risk that if the ambiguous sentence and its context required a more unusual (rare) interpretation, then participants might give the more common interpretation, not because of a problem in integrating linguistic information, but because of a problem in revising their initial interpretation. In other words, having the ambiguous sentence first might encourage acceptance of the most likely interpretation regardless of whether or not it is correct.

can interpret the ambiguous sentence context-appropriately. It does this by presenting a context-appropriate interpretation, a context-inappropriate interpretation and an erroneous interpretation. The participant is required to select the answer which best fits the meaning of the pair of sentences they heard. There is also a control task which requires participants to repeat two of the sentence pairs which required rare interpretations.

This test provides a quantitative measure of how well individuals can integrate auditorily presented linguistic information to establish meaning. If individuals with either autism or Asperger syndrome have weak central coherence, then the central coherence hypothesis of Frith (1989) would predict that they would be less proficient at integrating sentences to derive meaning. This should be evident on those ambiguous sentences requiring a more rare interpretation, since these require integrating the unusual context. In the absence of integrating prior context, so that the ambiguous sentence is processed in isolation, access to the meaning of the ambiguous sentence would be based on generating the most common meaning of the sentence, which would be contextually inappropriate when rare interpretations are required. The central coherence hypothesis would not, however, predict a problem with the more common interpretations of ambiguous sentences, since these do not have to be integrated with their prior context, and neither would it predict a problem in recalling the memory control sentences used in the test. Thus, the prediction was made that individuals with autism or Asperger syndrome would be less proficient at interpreting the ambiguous sentences requiring a rare interpretation, whilst performing normally on those requiring a more common interpretation and on the memory control task.

4.1. Participants

The same participants took part here as those who took part in the first experiment.

4.2. Materials

A tape recorder was used to present pairs of sentences. There were two pairs of trial sentences and 24 pairs of test sentences. The pairs of sentences were statements. The second sentence was always ambiguous, but it could be disambiguated by the first sentence which formed a biasing context.

There were two trial cards and 24 test cards which were to accompany each of the pairs of sentences presented on the tape recorder. Each card contained a question about an ambiguous sentence, along with three possible answers, only one of which was correct both for its corresponding ambiguous sentence and its biasing context. Each question with its three choices appeared on a separate card.

The 24 pairs of test sentences were made up of 12 ambiguous sentences, each of which were presented twice, but both of which were biased by two differing sentence contexts. One of the biasing contexts gave rise to a common interpretation of the ambiguous sentence and the other to a rare interpretation. Common and rare interpretations were based on frequency data collected from 100 clinically normal individuals. These individuals were presented with an extensive set of ambiguous sentences, and were requested to write down their first and then second interpretation of what the sentence meant to them. From these sentences, a subset of ambiguous sentences were chosen for use in this test. The sentences that were chosen were the ones that were most likely to be interpreted in one way rather than another. The criteria for selecting ambiguous sentences was that one interpretation (the common one) had to be at least three times more likely to be made than the other (rare) interpretation. The ratio was generally very much higher. Thus the ratio of common to rare interpretations for lexically ambiguous sentences ranges from 5 to 24 common interpretations for every one rare interpretation. The ratio of common to rare interpretations for syntactically ambiguous sentences ranges from 3 to 9 common interpretations for every one rare interpretation. Ambiguous sentences which were more likely to be interpreted in one way rather than another were ideal for the test, because when a rare interpretation was required, participants had to be particularly sensitive to the prior context in order to make the correct interpretation of the ambiguous sentence.

For the ambiguous sentences chosen, the clinically normal individuals' first and second most common interpretations formed two of the three choices given on the test cards. The third choice on the test card was always meaningless for both the ambiguous sentence and its biasing context. However, it did contain a word which was the key noun in its accompanying ambiguous sentence and was in similar syntactic form. This third sentence acted as a control for fatigue and for problems with motivation and attention. Half of the ambiguous sentences were lexically ambiguous and half were syntactically ambiguous. The mean length of the lexically ambiguous sentences was 6.65 words and the mean length of the syntactically ambiguous sentences was 8.05 words.

There were four types of sentence: lexically ambiguous sentences with a rare interpretation, lexically ambiguous sentences with a common interpretation, syntactically ambiguous sentences with a rare interpretation, and syntactically ambiguous sentences with a common interpretation. Each card was assigned a number and letter. For ambiguous sentences requiring a common interpretation, these were labelled on the reverse of the card C, and for ambiguous sentences requiring a rare interpretation, these were labelled on the reverse of the card R. Each ambiguous sentence was also assigned a number from 1 to 12 to accompany the letters C and R. Each number appeared twice so that the same ambiguous sentence shared the same number. The labels on the reverse of each card were to facilitate scoring.

A stopwatch was used to time how long it took the participant to make their choice. The position of the correct sentence and the distractor sentences was counterbalanced, so that the correct and incorrect interpretations of each ambiguous sentence appeared in each of the three possible positions eight times (twice in each of the four conditions).

The sentences used in this experiment were behavioural rather than mentalistic in content. It was found that the sentences were relatively easy to remember and comprehend. Examples of the stimuli used in this experiment are shown in Appendix A.

4.3. Procedure

The experimenter and participant were seated at a table in a room free from distractions. The set of sentence cards were placed face downwards on the table.

4.3.1. The trial items

There were two trial sentence cards which were presented in a fixed order. The experimenter told the participants they were going to hear some pairs of sentences. They were instructed to listen very carefully, as straight after they had heard them, they would be required to choose which one of three short sentences best fitted the meaning of the pair of sentences they heard. The experimenter played the first pair of trial sentences and then placed face up on the table its corresponding sentence choice card. Participants were encouraged to identify what the correct (interpretation) was. Most individuals could do so. The experimenter then played the second trial sentence and again individuals were encouraged to identify what the correct (interpretation) was. One individual with autism required some help on both the trial items. However, the experimenter made sure that all the participants understood why a particular sentence was the correct interpretation before introducing the test items.

4.3.2. The test items

The participants were then told 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 select what the correct sentence was. They were instructed that there was no time limit, but they were to say whether the top, middle or bottom sentence was correct. Participants were told to listen very carefully as they would only get to hear the pairs of sentences once. Because this was a forced choice test, individuals were encouraged to make a guess even if they did not know what the correct sentence was. Also, they were warned that at some point in the task they would on two different occasions be required to repeat back the pair of sentences which they had just heard, but that they would not be given any prior warning as to which pairs of sentences they would be required to recall. All participants fully understood what was expected of them, so the experimenter proceeded with the test sentences. The four types of sentence were mixed together and one of two different orders were presented to each individual in order to control for order effects. Participants were randomly assigned to each order and these two orders were counterbalanced across groups.

The two pairs of sentences which participants had to repeat back during the task were of different types: one was lexically ambiguous, the other was syntactically ambiguous. However they both had a biasing context, which would, by normal interpreters of the sentences (those that provided the frequency data), lead to a more rare type of interpretation. This acted as a control for working memory problems (as well as controlling for fatigue, motivation and attentional factors), as participants were requested to repeat the two sentences they heard after they had made their choice. The total number of words in each type of sentence pair to be recalled, i.e. whether lexically or syntactically ambiguous, had their lengths equal to the mean length of their type of sentence pair (lexical or syntactic) in the task as a whole. Irrespective of which of the two orders of ambiguous sentences an individual received, participants were requested to repeat back the same two pairs of sentences.

4.3.3. Scoring

To record response times, the stopwatch was started as soon as the card was turned over and placed down on the table. As soon as the participant reported whether the correct sentence was on the top, middle or bottom, the experimenter stopped the stopwatch. All response times were scored in s and 100th of a s, because some participants responded quite rapidly. Recording response times was aimed at assessing whether any difficulty shown by the autism group was due to finding auditory integration difficult (reflected in longer response times) or impulsiveness (presumed to result in faster response times, particularly on the failed items).

Type of choice made was also noted. There were three types of choice. One was the correct interpretation for the ambiguous sentence with its context (contextually-appropriate); the second was a possible interpretation for the ambiguous sentence when heard in isolation of its context, but which was a wrong interpretation for the ambiguous sentence when biased with its context (contextually-inappropriate errors); and the third type of choice was totally wrong. That is, it contained a word which was the key noun in its accompanying ambiguous sentence and was in similar syntactic form, but made little sense for either the ambiguous sentence or its biasing context. Such errors were likely to result from problems with attention, motivation or fatigue and as such were called ‘attention deficit’ errors. Therefore, there was only one correct answer, but a participant could make two types of error. The accuracy score for each participant was the number of contextually correct interpretations.

Performance on the memory control task was assessed in terms of the number of pairs of sentences recalled correctly (or correctly paraphrased so as to preserve the original meaning of the test sentences). The recall attempts (whether paraphrases or not) made by nine of the participants in each group 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 two pairs of sentences was 98%.

4.4. Results

The measure of whether answers were selected correctly for the ambiguous sentences biasing context constituted an accuracy score, which ranged from 0 to 6 for each of the four conditions. The mean accuracy scores for the three groups, by condition-type, are illustrated in Table 4.

4.4.1. Accuracy in choosing context sensitive answers

A three-way repeated measures ANOVA was performed on the mean accuracy scores. This ANOVA had one between-participant variable of group, and two within participant variables of type (lexical and syntactic) and frequency (common and

Table 4
Number of ambiguous sentences interpreted context appropriately. Data are mean ± SD with ranges in parentheses.

Participant group (n = 17)	Common interpretations		Rare interpretations	
	Lexical	Syntactic	Lexical	Syntactic
Normal	5.41 ± 0.71 (4–6)	5.29 ± 0.69 (4–6)	5.41 ± 0.71 (4–6)	5.53 ± 0.80 (4–6)
Autism	5.29 ± 0.69 (4–6)	4.94 ± 0.66 (4–6)	3.94 ± 1.20 (2–6)	3.82 ± 1.19 (2–6)
Asperger	5.41 ± 0.87 (3–6)	5.18 ± 0.64 (4–6)	4.65 ± 1.00 (3–6)	4.71 ± 1.11 (3–6)

rare). This analysis revealed a significant main effect of group ($F(2,48) = 8.91$, $P = 0.001$). Further exploration of this effect with the Newman–Keuls test showed that the autism group was significantly less accurate than the normal and Asperger groups ($P < 0.01$; $P < 0.05$, respectively). Although the Asperger group did not differ from the normal group, there was a trend for the Asperger group to be less accurate (see Table 4).

There was a non-significant effect of type ($F(1,48) = 1.52$, $P = 0.22$), which indicates that the type of stimuli, whether lexically or syntactically ambiguous, did not have a significant effect on mean accuracy scores.

The first-order interaction of group by type was non-significant ($F(2,48) = 0.62$, $P = 0.54$). Therefore, none of the groups were differentially affected by the type of ambiguous sentence. The first-order interaction of frequency by type was also non-significant ($F(1,48) = 1.48$, $P = 0.23$). Therefore, the frequency of the interpretation did not have an effect on the type of ambiguous sentence, and vice-versa.

The analysis revealed a significant main effect of frequency ($F(1,48) = 31.40$, $P < 0.001$), which the means suggest can only really be interpreted as the result of its higher-order interaction with group (group by frequency: $F(2,48) = 14.35$, $P < 0.001$). Simple effects were examined which compared the two frequency conditions for each group. Analysis of simple effects showed the effect of frequency to be significant for the autism and Asperger groups (F autism $(1,48) = 47.73$, $P < 0.001$; F Asperger $(1,48) = 11.93$, $P = 0.001$), but not the normal control group ($F(1,48) = 0.43$, $P = 0.51$). Observation of the means (see Table 5 and Fig. 1) show the autism and Asperger groups to be significantly worse on the rare conditions relative to their own performance on the common conditions, whereas the normal control group showed no differential pattern of performance on the two conditions.

Table 5
Types of error on the ambiguous sentences requiring a rare interpretation. Data are mean ± SD with ranges in parentheses.

Participant group (n = 17)	Contextually inappropriate	Attention deficit
Normal	0.94 ± 1.20 (0–3)	0.12 ± 0.33 (0–1)
Autism	3.82 ± 2.04 (0–7)	0.41 ± 0.62 (0–2)
Asperger	2.29 ± 1.76 (0–6)	0.35 ± 0.61 (0–2)

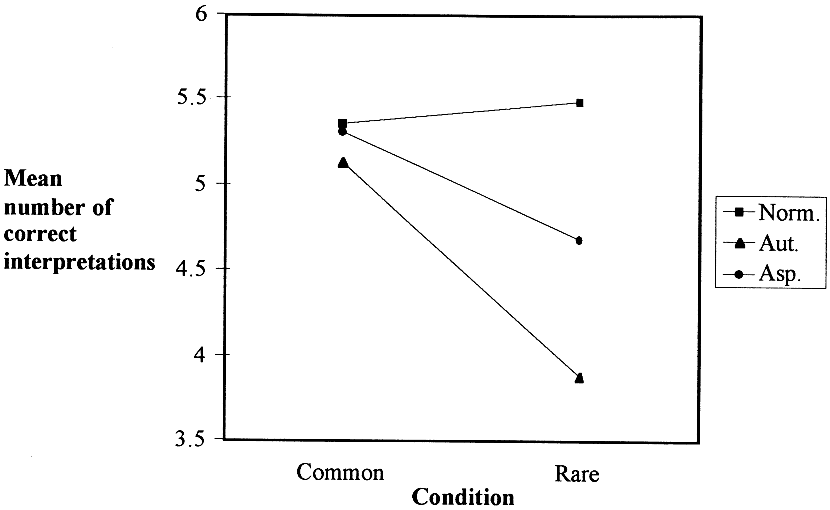

Fig. 1. Effect of frequency of ambiguous sentence on correct interpretation of the sentence.

Given that the group main effect was significant, the group by frequency interaction was investigated further to see if there were different group effects for the two frequency conditions. Simple effects were examined which compared the different groups on each of the frequency conditions. As predicted, this showed the effect of group to be significant only for the rare condition ($F(2,48) = 13.73$, $P < 0.001$). The source of this interaction was investigated further using t-tests. The data was normally distributed and the variances were approximately equal, so a pooled variance estimate was used. To reduce the family-wise error rate, only pre-planned comparisons were explored. To further control the family-wise error rate, the two rare conditions were combined, as keeping these two conditions separate was not particularly meaningful, especially since the type of sentence was found not to interact with frequency. Planned contrasts of the cell means indicated that the rare mean for the autism and Asperger groups was significantly different to that of the normal group (t autism (48) = 5.24, $P < 0.001$; t Asperger (48) = 2.62, $P < 0.05$) and that on the rare items the clinical groups themselves differed ($t(48) = 2.62$, $P < 0.05$). Observation of the means (see Table 4 and Fig. 1) show the clinical groups to be significantly worse in choosing the interpretation which was correct for the context, but only when the context required a more rare interpretation. Furthermore, the autism group's performance on the rare items was significantly worse than the Asperger group's performance on the rare items.

None of the groups revealed a frequency by type interaction since the second-order interaction of group by frequency by type was non-significant ($F(2,48) = 0.01$, $P = 0.99$).

4.4.2. Analysis of error types

Participants could make two types of error. One type was a contextually incorrect interpretation of the ambiguous sentence (contextually-inappropriate errors). These interpretations failed to take account of the ambiguous sentence's prior context. The other type of error could not really be considered to be an interpretation at all. These answers were incorrect irrespective of the ambiguous sentence and irrespective of its biasing context (attention deficit errors). Examining just the sentences requiring a rare interpretation (in order to reduce the family-wise error rate), the frequencies of the two error types were calculated (see Table 5). The data was approximately normally distributed for each group and there were no obvious outliers, so two one-factor ANOVAs were performed, one on each error type. The first ANOVA revealed that the groups did not differ in the number of attention deficit errors ($F(2,48) = 1.44$, $P = 0.25$). However, a second one-factor ANOVA on the contextually-inappropriate errors, did reveal a significant difference between groups ($F(2,48) = 12.21$, $P < 0.001$). This group effect was investigated further using *t*-tests. Planned contrasts indicated that the autism and Asperger groups chose significantly more contextually-inappropriate errors (t autism $(48) = 4.94$, $P < 0.001$; t Asperger $(48) = 2.32$, $P < 0.05$) and that the autism and Asperger groups themselves differed, the autism group making significantly more of this type of error ($t(48) = 2.62$, $P < 0.05$). Whether one looks at the mean number of contextually-inappropriate errors or the mean number of contextually-appropriate responses, the data analysis yielded the same results.

4.4.3. Memory control

The evidence from the memory control task did not support there being a memory problem (see Table 6), since a 3×2 Pearson Chi-square test did not reveal a difference between groups on the number of pairs of sentences which were recalled correctly ($\chi^2(2) = 2.99$, $P = 0.22$).

4.4.4. Response time

The raw data was transformed into logarithms (base 10) to reduce skewness and stabilise the variances which were unequal due to outliers and extreme values. Then a three-factor repeated measures ANOVA was performed on the mean response times for the three groups. Again, this ANOVA had one between-participant variable of group, and two within participant variables of type (lexical and syntactic) and frequency (common and rare). This analysis revealed a significant main effect of Group ($F(2,48) = 3.91$, $P < 0.05$). Further exploration of this effect with the Newman–Keuls test showed that the autism group took significantly longer to make their response than the normal and Asperger groups ($P < 0.05$).

There was a significant effect of type ($F(1,48) = 44.84$, $P < 0.001$), which indi-

Table 6
Pairs of sentences recalled

Participant group ($n = 17$)	Recalled correctly
Normal	17
Autism	23
Asperger	23

cates that the type of stimuli whether lexically or syntactically ambiguous had a significant effect on mean response times. Observation of the means show the syntactic conditions' responses are significantly slower than the lexical conditions' responses (see Table 7).

The analysis also revealed a significant main effect of frequency ($F(1,48) = 5.53$, $P < 0.05$), which indicates that response time varied according to whether the ambiguous sentence and its context required a common or rare interpretation. This effect was due to participants taking significantly longer when the task required a more rare interpretation, rather than when the task required a more common interpretation. Observation of the means in Table 7 shows this main effect to be due mainly to the two clinical groups, who were slower on the sentences requiring a more rare interpretation.

The first-order interaction of group by type was not significant ($F(2,48) = 1.13$, $P = 0.33$). Therefore, none of the groups were differentially affected by the type of sentence. The first-order interaction of frequency by type just reached significance ($F(1,48) = 4.34$, $P < 0.05$). Therefore, the frequency of the interpretation interacted in some way with the type of ambiguous sentence. Given that the frequency main effect was significant, it was useful to see whether there was a frequency effect for the type of stimuli. Simple effects were examined which compared frequency for each type of stimuli. Analysis of simple effects showed the effect of frequency to be significant for the lexical condition ($F(1,48) = 8.58$, $P < 0.01$), but not the syntactic condition ($F(1,48) = 0.00$, $P = 1.00$). Observation of the means show that rare lexical interpretations are responded to significantly slower than the common lexical interpretations (see Table 7).

Given that the type main effect was significant, then the frequency by type interaction was investigated further to see if there was a type effect for the two frequency conditions. To examine whether this type effect applied to one or both of the frequency conditions, simple effects were examined which compared the type of stimuli in each of the frequency conditions. Analysis of simple effects showed the effect of type to be significant for both the common and rare conditions (F common $(1,48) = 41.02$, $P < 0.001$; F rare $(1,48) = 6.43$, $P < 0.05$). Observation of the means (see Table 7) show the syntactic items to be responded to significantly slower than the lexical items, whatever the frequency of the stimuli.

Table 7
Ambiguous sentences response times. Data are mean ± SD with ranges in parentheses.

Participant group ($n = 17$)	Common interpretations		Rare interpretations	
	Lexical	Syntactic	Lexical	Syntactic
Normal	4.06 ± 1.57 (2.00–10.00)	5.92 ± 1.39 (3.16–10.00)	4.45 ± 1.65 (1.58–12.59)	5.40 ± 1.63 (2.51–12.59)
Autism	5.43 ± 1.51 (3.16–10.00)	7.52 ± 1.53 (3.98–15.85)	6.89 ± 1.64 (3.16–15.85)	8.26 ± 1.67 (3.16–19.95)
Asperger	5.18 ± 1.31 (3.16–7.94)	6.70 ± 1.33 (3.98–12.59)	6.27 ± 1.65 (2.00–15.85)	6.68 ± 1.60 (2.51–12.59)

The first-order interaction of group by frequency was not significant ($F(2,48) = 1.70$, $P = 0.19$). However, since one of the reasons for examining response times was to see whether the clinical groups were slower on the rare items, simple effects were used to examine whether groups differed in their response time on the rare items. This showed the effect of group to be significant on the rare condition ($F(2,48) = 4.01$, $P < 0.05$). Performance on the rare condition was investigated further using *t*-tests. The data was normally distributed and the variances were approximately equal so a pooled variance estimate was used. To reduce the family-wise error rate, only pre-planned comparisons were explored. To further control the family-wise error rate, the two rare conditions were combined, as keeping these two conditions separate was not particularly meaningful. Planned contrasts of the cell means indicated that the mean response time for the autism group's rare condition was significantly slower than that of the normal group ($t(48) = 2.79$, $P < 0.01$) and that the mean response times for the Asperger group's rare condition approached significance when compared with the normal control group ($t(48) = 1.81$, $P = 0.08$). The clinical groups' mean response time for the rare items did not differ ($t(48) = 0.98$, $P = 0.33$). Observation of the means show the clinical groups to be slower in choosing the interpretation which was correct for the context (see Table 7).

None of the groups revealed a frequency by type interaction since the second-order interaction of group by frequency by type was non-significant ($F(2,48) = 0.04$, $P = 0.97$).

4.5. Discussion

The clinical groups performed normally on the common condition – ambiguous sentences which could be interpreted correctly irrespective of their context, by just selecting the most common interpretation. However, the clinical groups performed significantly less well than their normal control group on the rare condition – ambiguous sentences which could only be interpreted correctly by being sensitive to their prior context – processing for higher-level meaning. This was reflected not only in their accuracy scores on this condition, which were significantly lower for both clinical groups, but in their response times on this condition, which were longer for both groups and significantly longer for the autism group. Once again, this result is exactly as predicted by central coherence theory. The clinical groups themselves differed, the autism group being significantly less able to select the context-appropriate interpretation of the ambiguous sentence in the rare condition than the Asperger group, although these groups did not differ in their response times on this condition.

When looking at performance within groups, it was found that the normal control participants showed no differential pattern of performance on the accuracy of their common and rare conditions, whereas the autism and Asperger groups performed significantly less well on the rare condition relative to their own performance on the common condition.

The analysis also revealed that response time varied according to whether the ambiguous sentence and its context required a common or rare interpretation. This difference was due to individuals taking significantly longer when the task required a more rare interpretation, rather than when the task required a more common interpretation. This is what one would expect, as individuals would tend to take longer when the interpretation required was not one they would normally expect. However, as mentioned previously, this effect was due to the two clinical groups being slower on the sentences requiring a more rare interpretation. Like the rare and common items, the syntactic and lexical items evidenced significant response time differences. Thus, the syntactic items were responded to significantly slower than the lexical items, whatever the frequency of the stimuli. Intuitively this is what one would expect, since the syntactic items had a mean length of sentence which was longer than that of the lexical items and they were also grammatically more complex. However, none of the groups' accuracy in selecting the context-appropriate interpretation of the ambiguous sentences were differentially affected by the type (lexical or syntactic) of ambiguous sentence.

The rare-lexical condition was responded to significantly slower than the common-lexical condition. There was no response time difference between the rare-syntactic and common-syntactic conditions. There are two reasons for these findings. Firstly, as already mentioned, one would expect individuals to take longer when the interpretation required was not one that they would normally expect, although the difference was mainly due to the clinical groups being much slower. Secondly, the fact that there was no significant difference in response times for the common and rare syntactic items is probably due to there not being such a clear cut difference between what was a common and rare interpretation for each of the syntactically ambiguous sentences, at least not in the way there was for lexically ambiguous sentences (see the ratios of common to rare interpretations in Section 4.2).

Further support for the idea that the problem for the clinical groups was one of processing for higher-level meaning can be seen on the working memory control task. This task required participants to hold the sentences (or the meaning of them) in memory whilst they selected the interpretation they believed to be correct, and only after this were they required to recall them. The clinical groups tended to recall the exact wording of these sentences, and this was even the case when they failed to integrate the sentences meaningfully. Their good performance on the memory task, in contrast to their poor interpretation of context, suggests they pay preferential attention to surface form rather than the meaning. This observation for the clinical groups did not seem to apply to the normal group, at least not to the same extent. For most normal individuals, the recall attempts were paraphrases of the original sentences, i.e. the meaning was preserved while the exact details faded. Moreover, a minority of the normal individuals who gave the context-appropriate interpretation of the ambiguous sentence then could not recall the sentences. It is as if for these individuals, the meaning remained long enough for them to make their choice, but then faded.

The findings suggest that on the rare condition, the clinical groups had greater difficulty in selecting the context-appropriate interpretation. Given their normal performance on the common condition, which is in stark contrast to their relatively impaired performance on the rare condition, this suggests that the failure on the rare condition might be due to a problem in interpreting these sentences on the basis of their context. That is, the clinical groups were less efficient at integrating the ambiguous sentence with its prior context. This impairment in integration is reminiscent of the findings from Experiment 2, where the clinical groups were significantly less proficient at integrating sentences so as to select the coherent inference. However, unlike the previous experiment, this difficulty is also evident in the domain of auditorily presented information.

Data from this experiment bear on questions about how contextual ambiguity is resolved (e.g. Fodor and Garrett, 1967; Carey et al., 1970). Discussing how ambiguity is resolved is beyond the scope of this paper. However, it is generally accepted that ambiguity is resolved by appealing to the context (Becker, 1980, 1985; Forster, 1981; Onifer and Swinney, 1981; Stanovich and West, 1981, 1983; Simpson, 1984; Byrd, 1988). Thus, the clinical groups who had the greater difficulty in using context, made significantly more contextually-inappropriate errors than the normal control group, while none of the groups differed in terms of the number of attention deficit errors. Also of interest, is the fact that the autism group made significantly more contextually-inappropriate errors than the Asperger group.

The results indicate that it is the need to integrate the ambiguous sentence with its context that accounts for the clinical groups' difficulties. Furthermore, this deficit was found to be universal to both clinical groups. The finding that the autism group was significantly more impaired than the Asperger group is reminiscent of their comparatively greater failure to select the most coherent inference in Experiment 2. However, whereas Experiment 2 demonstrated a deficit with self-read material, Experiment 3 demonstrated a deficit with auditory material. The importance of this experiment is that not only does it provide evidence of a deficit in the auditory domain, but it at the same time clarifies the finding from Experiment 2, by demonstrating a deficit in achieving local coherence with non-mentalistic and non-script material.

Does the Ambiguous Sentence (AS) task, the Local Coherence Inference (LCI) task and the Homograph (Hom) task correlate with one another?

Since all three experiments were designed to tap weak central coherence, it was useful to see whether performance on the critical conditions correlated with one another. Pearson correlations found that the tasks were correlated for all three groups ($P < 0.05$: HFA: Hom × LCI = 0.61; LCI × AS = 0.73; Hom × AS = 0.52; Asp: Hom × LCI = 0.61; LCI × AS = 0.66; Hom × AS = 0.53; N: Hom × LCI = 0.59; LCI × AS = 0.60; Hom × AS = 0.50). Pearson correlations were also used in order to establish whether performance was correlated with intelligence (FSIQ, VIQ and PIQ). Thus for the HFA group, performance on the LCI and AS tasks correlated with all IQ measures ($P < 0.05$), but there were no significant correlations on the homograph test ($P \geq 0.06$). Similarly, for the Asperger group, performance on the LCI and AS tasks broadly correlated with IQ ($P < 0.05$ for FSIQ and VIQ; $P = 0.06$ for PIQ), but again there were no significant correlations on the homograph test ($P \geq 0.06$). For the normal control group, there were no significant correlations with any of the IQ measures ($P > 0.05$). Since IQ played a role (for the clinical groups on two out of the three tasks), the IQ measures were partialled out and performance on the three tasks were again correlated. Pearson correlations found that the correlations on the Hom × LCI and LCI × AS tasks were upheld for all three groups ($P < 0.05$), but the correlation on the Hom × AS tasks disappeared for all three groups ($P > 0.05$).

5. General discussion

Prutting (1982) noted that a universal feature of language is that it is contextual. This paper sought to assess the contextual aspects of language by testing the prediction from the central coherence theory that individuals with an autism spectrum condition will be impaired in their ability to achieve local coherence. The three experiments suggest that the clinical groups have difficulty in achieving local coherence, consistent with Frith's theoretically-based predictions. This difficulty is unlikely to be due to problems with motivation. Thus, one would have expected motivational differences to be evident on the common pronunciations (homographs) and common interpretations (ambiguous sentences) as well as on the rare pronunciations and rare interpretations, yet this was not the case. Furthermore, one would expect motivational differences to be evident on various types of control task, and yet this was not the case – the clinical groups performed normally on the sentence recall task (Experiment 3), and on the memory and comprehension questions (Experiment 2). These findings do not support a motivational deficit hypothesis. Neither do these findings support inattention and fatigue as playing a role, since these would predict a deficit on the memory and comprehension control tasks and there were no such deficits. Any suggestion that fatigue was the problem on the rare pronunciations and rare interpretations, would also predict that fatigue would be a problem on the common pronunciations and common interpretations, since the common conditions were always intermixed with the rare conditions. However, performance on the common conditions was found to be unimpaired. Moreover, on the Ambiguous Sentences test, neither of the clinical groups made more attention deficit errors, which one would have expected if there had been problems of fatigue or problems with attention and motivation.

It might be argued that the clinical groups' difficulty with the three tests stem from executive problems (Russell, 1997). However, there are a number of findings which do not support this. Thus, the clinical groups are unlikely to have performed poorly simply because of motor or cognitive impulsivity, since they tended to respond slower on items which they failed or had difficulty with, and had on the rare conditions of the Ambiguous Sentences test prolonged response times overall. Neither was there any evidence to support their responding perseveratively, i.e. always choosing the answer which was in a particular position on a stimulus card. Had they done so then they would have performed at chance in the Local Coherence Inferences and Ambiguous Sentences tests since each correct interpretation appeared in each possible position an equal number of times. Moreover, had they behaved perseveratively on the Ambiguous Sentence test, a deficit would be expected on the common condition as well as the rare condition, yet there was no such deficit. Similarly, it is unlikely that the clinical groups' difficulties were a secondary consequence of working memory difficulties, since on the Local Coherence Inferences test the stimuli remained in view, and on the Ambiguous Sentences test they performed well on the working memory control task. Nor was there any evidence of the clinical groups failing to monitor their reading. Thus on the homograph task the difficulty on the rare pronunciations did not seem to stem from a failure to make self-corrections. Across the four conditions, 88% of the adults with autism and 65% of the adults with Asperger syndrome who performed below ceiling and so could have corrected themselves, did so, compared to 76% of the normal adults.

It is equally unlikely that the group differences stemmed from comprehension problems, since each group made just one or two errors on the Local Coherence Inferences test comprehension question, and on the Ambiguous Sentence test the clinical participants had no problem in providing interpretations in the common condition. Furthermore, the ambiguous sentence scenarios were very simplistic, none were particularly social or mentalistic in nature, nor did they contain much in the way of emotional language or make demands on script knowledge. Moreover, the type of error made by the clinical groups on this task was suggestive of a contextual problem, rather than a comprehension problem, since they made significantly more of the contextually-inappropriate errors, but not more of the attention deficit errors.

Given the significance of the linguistic deficit observed in the clinical groups, and their normal performance on the Local Coherence Inferences test's comprehension question, one cannot help but question what determines these different results, i.e. why they perform normally on the comprehension question, but not on, for example, selecting the most coherent inference. However, it is clear that the comprehension question does not require one to integrate elements to the same extent as does selecting the coherent inference. Selecting the coherence inference is much more demanding. It is a complex process requiring the pulling together of information from disparate sources, so that the various elements of the text are related. Thus, details have to be woven together in order to establish the context. This may be quite distinct from just comprehending a given context. Furthermore, in order to construct a coherent representation, readers must identify concepts in each sentence that refer to the choices available. Very often the difference between the correct and incorrect inferences rests on the interpretation of an explicitly stated word or phrase. It is these factors which makes this test different from many other tests and straightforward reasoning tests, because of the emphasis it places on integration to establish high-level meaning. This difference between comprehension and weaving together details is very similar to a statement made by Rumsey and Hamburger (1988). These authors noted that the responses of high-functioning adults with autism to the verbal absurdities and problem situations from the Stanford–Binet intelligence test (Terman and Merrill, 1973; Thorndike et al., 1986) reflected comprehension of the linguistic aspects of these problems but failures in integration.

There are various linguistic arguments which might be put forward in order to try and account for the clinical groups' apparent deficit. The Local Coherence Inferences test can be construed as a linguistic reasoning test – reasoning to achieve the most coherent inference – but it is unlikely that a linguistic reasoning impairment per se could account for the clinical groups' difficulties since they have not been found to have a reasoning impairment (Raven et al., 1992; Scott and Baron-Cohen, 1996; Scott et al., 1999). It is also unlikely that the clinical groups' difficulty was due to poor semantic ability as such, since their comprehension of the meaning of single words is intact (Prior and Hall, 1979; Frith and Snowling, 1983; Minshew et al., 1995). Another argument pertains to reading skill and experience. These are not likely to be explanations for the deficit, because they were observed to read normally and had no difficulty in reading the pre-test list of homographs. Also, reading skill has been shown to be mental age appropriate or even superior (see Frith and Snowling, 1983). Furthermore, the vast majority of clinical subjects had successfully completed public examinations (and a proportion in each group possessed one or more university degrees) and all were of average or above average verbal and full-scale IQ, making it highly unlikely that the clinical groups' difficulty with these tasks was one of deficient language or intellectual ability per se. Crucially, the measures were designed to tap language in context, rather than language ability per se, and given that use of language in context is one of the key areas of difficulty for these individuals (Tager-Flusberg, 1981a) the finding is perhaps not that surprising.

Might it be the case that the clinical groups simply have poorer standards of coherence? Certainly individuals with a poorer standard of coherence would be less likely to select the most coherent inference on the Local Coherence Inferences test, because of being satisfied with a less coherent inference. However, response times did not support this, as one would have expected such an approach to result in faster not slower response times. Moreover, if their standard of coherence was the factor behind their poor ability, this would still beg the question as to what causes their poorer standard of coherence. Although standards of coherence is new to the study of psycholinguistics (see Van den Broek et al., 1995; Britton and Graesser, 1996), maintaining a poorer standard of coherence is believed to be a function of motivation, the purpose for reading, the time at one's disposal to read, reading experience and reading ability. Standards of coherence must be a consideration on any future tests which seek to assess linguistic coherence.

Despite methodological variations (i.e. integrating a homograph with its sentence context, integrating two sentences in order to disambiguate the second sentence, and selecting an inference which makes a situation and its outcome coherent) the results of these three experiments converge on the same basic conclusion: a difficulty in processing linguistic context so as to achieve local coherence. This is very encouraging evidence for a central coherence theory. Thus, the findings from the Local Coherence Inferences test and the Ambiguous Sentences test suggest that individuals with autism or Asperger syndrome have real difficulties in integrating linguistic material for meaning. A recent study which has also found a difficulty in linguistic integration looked at the ability of high-functioning adults with autism to judge whether an inference was correct or not (Ozonoff and Miller, 1996). These authors presented participants with two sentence scenarios and two inferences. One of the inferences was correct for the two sentences when processed together, and the other was incorrect for the two sentences, since it was based on only one of them. The adults with autism demonstrated difficulty in integrating information presented in the two sentences so as to make the judgement as to whether an inference was correct or not.

A primary activity during reading or listening is connecting the phrase or sentence that is currently read or heard with the contents of the immediately preceding sentence. If an individual is able to make such a connection, then local coherence can be said to have been achieved. This is consistent with what Haviland and Clark (1974) claimed when they proposed that each new sentence is comprehended with respect to the information contained in the previous sentence. The ability to make such connections and thus to disambiguate what is read and heard is vital for maintaining coherence. Information processing in real life almost always involves the interpretation of individual stimuli in terms of overall context. If these clinical groups have a decreased ability to interpret information in context, then comprehension and discourse coherence will suffer. It is clear that such difficulties would disrupt communication ability and it may partly explain their insensitivity to the pragmatic aspects of communication. It has become increasingly clear in recent years, that pragmatics, which is the use of language in context, is the area of language that is seriously impaired in autism (Tager-Flusberg, 1981a; Baron-Cohen, 1988; Tantam, 1991; Happé, 1993; Surian et al., 1996). Also, since much of learning involves the integration of stimuli, failure to connect stimuli meaningfully might be an important factor in the development of autism.

The autism group had greater difficulty integrating linguistic information in comparison to the Asperger group. This is not surprising since as a group these individuals tend to have a greater disability in childhood (and sometimes in adulthood) than those with Asperger syndrome. This may be empirical support at the cognitive level that autism is more severe than Asperger Syndrome. However, the finding that the Asperger group performed significantly below that of the normal group on all three tasks is quite striking, since individuals with Asperger syndrome do not have the (clinically significant) early language delay that characterises those with autism. This finding would seem to suggest that early language development per se cannot play a major role on tasks which assess linguistic coherence. Given that the deviant language which characterises individuals with Asperger syndrome does not become noticeable until after the first few years of life, when the use of language in context becomes more critical, it is interesting that the language tasks used here pick up a relatively pervasive problem in processing language in context.

The fact that weak central coherence was found with three quite diverse tasks suggests that weak central coherence might stem from a deficit in central thought processes. This being so, the three tasks should be positively and highly correlated for each of the three groups. This was found, and tends to suggest that the coherence results were not an artifact of any single method employed. It also suggests that weak central coherence might be a unitary force in these different tasks. Had the correlations not been significant or had they been lower, it would have begged the question that there might be different types of coherence, only some of which are impaired in autism spectrum disorder. However, for the clinical groups at least, these correlations might have been in part due to performance being mediated by intelligence since there were correlations with intelligence (FSIQ, VIQ and PIQ). This was the case on the LCI and AS tasks, although not for the homograph test. There was, however, a trend for the clinical groups' performance to be correlated with IQ on the homograph task, but the failure to find significant correlations with this task may reflect the fact that even for the clinical groups the homograph task was relatively easy, regardless of intelligence level. Not surprisingly and possibly for the same reason, the normal control group also failed to show any correlation with IQ on this same task. However, for the normal control group there was also a failure to find significant correlations with IQ for the other two tasks (LCI and AS), although there was a trend for performance to be related to IQ. This could suggest that the significant correlations found with both the clinical groups on the LCI and AS tasks might be due to them using a different strategy to their normal control group. This is an area for future research. It might also suggest that central coherence rather than just intelligence was mediating the performance of the normal group.

Since it was clear that IQ played some role for all three groups, this was partialled out and the correlations repeated. The correlations were upheld on the Hom $\times$ LCI and LCI $\times$ AS tasks, but the correlation on the Hom $\times$ AS tasks disappeared for all three groups. Given that the Hom $\times$ AS correlation was no longer significant, it is useful to explore the nature of the three tasks. For the Hom and LCI tasks, both are reading tasks, although in the former spontaneous reading for meaning is tested, whilst in the latter it is enforced reading for meaning. For the LCI and AS tasks, both require integrating information for meaning, whilst one assesses reading and the other listening to spoken material. For the Hom and AS tasks, there is less in common, the homograph task is a reading task which assesses spontaneous reading behaviour, whereas the AS task is a listening task which assesses enforced listening. It might be that the earlier significant correlations found on these tasks do not withstand partialling out IQ because they have less in common. This again is an area for future research.

In conclusion, the results from the three experiments provide strong support for a 'central coherence' account of autism deficits (Frith, 1989). Frith's central coherence hypothesis predicts deficits across the three tasks, since all require the derivation of meaning. The findings also demonstrate the predicted deficit across domains, i.e. with self-read and auditory material. However, despite the significant findings, the magnitude of the critical differences were rather small, therefore it might be more accurate to talk in terms of a relative inefficiency rather than a major impairment. But what is noticeable, is that across the three tests, the clinical groups' scores on the critical conditions did not overlap very greatly with their normal control group. Moreover, the finding of a preference not to fully process for meaning, as well as a deficit in processing for meaning, suggests that individuals with an autism spectrum disorder have to make a greater than normal effort to process for meaning, with the result that they tend not to fully process for meaning unless requested to do so or unless they make a conscious decision to do so. Furthermore, as one would expect, the finding of a deficit in integrating linguistic information at the local level must also occur at the global level and there is recent evidence which demonstrates this (Jolliffe and Baron-Cohen, 1997). However, the conclusion which we draw from these experiments is that high-functioning individuals with autism or Asperger syndrome are inefficient in achieving local coherence.

Acknowledgements

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 grateful to Stephen Monsell and Ian Nimmo-Smith for advice on methodology. The first author was supported by a 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.

Appendix A. Examples of test sentences

The Homograph test

Pre-test items

Bow lead tear row read bat ball glasses letter palm nail bank light

Examples of test items

Rare pronunciation: before context

(1) It was lead in the box that made it so heavy.
(2) The man had a second row with his wife the day after.

Rare pronunciation: after context

(1) The scrap metal man first took the copper and iron and then he took the lead.
(2) The brothers started shouting. Dick left because he didn't want to be involved in a row.

Frequent pronunciation: before context

(1) It was the lead guitarist that sang at the concert.
(2) The man had a second row seat in the cinema.

Frequent pronunciation: after context

(1) Mary wanted to take the dog for a walk, so she went to the cupboard and took the lead.
(2) Everyone who wanted to see the new film had to stand in a row.

Local Coherence Inferences

Trial item

George left his bath water running.

George cleared up the mess in the bathroom.
George cleared up the mess in the bathroom because:

the bath had overflowed
his brother had left it untidy
the workman hadn't cleared up his mess

Examples of test items
Script
Albert said he wouldn't return to the restaurant.
He left without giving a tip.
Albert didn't leave a tip because:

he only had enough money to pay for the meal
he was dissatisfied with the service
the restaurant was closed when he arrived

Mentalistic
Ann took the bus into town to buy her mother a dress.
Ann visited six shops and travelled home feeling miserable.
Ann felt miserable because:

she couldn't afford a dress her mother would have liked
she couldn't find her mother's size of dress
she had lost all her money in one of the shops

Mentalistic/script
Michelle waited for over an hour for Jim outside MacDonalds.
Michelle ran home and refused to take any phone calls.
Michelle didn't want to talk on the phone because:

she was angry with Jim and wasn't going to hear his excuses
she went home to ask her sister where Jim might be
she was too upset to speak in case something had happened to Jim

Non-mentalistic/non-script
Dawn hurried to the chemist.
Two hours later Dawn noticed her headache had gone away.
Dawn's headache had gone away because:

she bought some painkillers
she had some fresh air on her walk
she let nature take its course

Non-mentalistic/non-script
From a distance the campers watched as the great mountain rumbled.
The TV showed fields of black tree stumps instead of a forest.
The forest turned into black tree stumps because:

vandals had set the forest on fire

a volcano erupted and lava poured into the forest
a bomb had exploded, setting the forest on fire

Control task

Philip went to the horse fair to look for a very good horse. It was not long before he found a horse which he could take around the shows. Philip bought the horse and took him home and kept it in his field. He spent over a year training his horse until they were ready to tour the shows. He then went to London to buy a new set of riding clothes. When he got back to the field his horse was grazing, he whistled and called his horse by name. But Philip looked so different in his new clothes that his horse just carried on grazing. Philip called again. The horse still wouldn't come over. So Philip went and put on his old riding clothes. When he returned to the field, his horse immediately galloped over. Philip patted and then groomed the animal for they were going to a show. (*Reading speed*).

Did Philip buy his horse from a horse fair? (*Memory*).

Did the horse come over when Philip was dressed in his new riding clothes? (*Comprehension*).

The Ambiguous Sentence test

Trial items

(1) John went to his art class.
He drew a gun.
What did John do?

pull out a gun
draw a picture of a gun
shoot from a gun

(2) The lady was waiting for the postman to call.
She looked up the street.
What did the lady do?

look along the street
look for the street on a map
look for a shop in the street

Examples of test items

Rare interpretation: lexically ambiguous

(1) The boiler house was very noisy.
The roar of the fans disturbed the team.
What happened?

football fans disturbed the team
football fans helped the team
cooling fans disturbed the team

(2) Clare was robbed as she walked along by some water.
The bank was the scene of the robbery.
Where was the robbery?

on a river bank
in a bank
in the village bank

Common interpretation: lexically ambiguous
(1) The crowd was very noisy.
The roar of the fans disturbed the team.
What happened?

football fans disturbed the team
cooling fans disturbed the team
football fans helped the team

(2) A man pointed a gun at the cashier.
The bank was the scene of the robbery.
Where was the robbery?

in the village bank
on a river bank
in a bank

Rare interpretation: syntactically ambiguous
(1) Mrs Jackson was fond of animals.
The shooting of the hunters upset the old lady.
What had upset the old lady?

the hunters dying
the forest being on fire
the hunters killing badgers

(2) The woman liked to keep her house tidy.
She said that visiting relatives can be a nuisance.
What did she mean?

having relatives visit was a nuisance
having relatives was a nuisance
having to visit relatives was a nuisance

Common interpretation: syntactically ambiguous
(1) Mrs Jackson used to visit her friends when passing through the forest.
The shooting of the hunters upset the old lady.
What had upset the old lady?

the hunters killing badgers
the hunters dying
the forest being on fire

(2) The woman hated travelling.
She said that visiting relatives can be a nuisance.
What did the lady mean?

having relatives was a nuisance
having relatives visit was a nuisance
having to visit relatives was a nuisance

Control task:
Rare interpretation: lexically ambiguous sentence

The boiler house was very noisy.
The roar of the fans disturbed the team.

Rare interpretation: syntactically ambiguous sentence

The woman liked to keep her house tidy.
She said that visiting relatives can be a nuisance.

References

Abelson, R.P., 1981. Psychological status of the script concept. American Psychologist 36, 715–729.

American Psychiatric Association, 1987. Diagnostic and Statistical Manual of Mental Disorders (DSM-IIIR), 3rd (revised) edn. American Psychiatric Association, Washington DC.

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), 4th edn. American Psychiatric Association, Washington DC.

Anderson, A., Garrod, S.C., Sanford, A.J., 1983. The accessibility of pronominal antecedents as a function of episode shifts in narrative text. Quarterly Journal of Experimental Psychology 35A, 427–440.

Baron-Cohen, S., 1988. Social and pragmatic deficits in autism: cognitive or affective? Journal of Autism and Developmental Disorders 18, 379–402.

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

Bartolucci, G., Pierce, S., Streiner, D., Eppel, D., 1976. Phonological investigation of verbal autistic and mentally retarded children. Journal of Autism and Childhood Schizophrenia 6, 303–316.

Becker, C.A., 1980. Semantic context effects in visual and word recognition: an analysis of semantic strategies. Memory and Cognition 8, 493–512.

Becker, C.A., 1985. What do we really know about semantic context effects during reading? In: Besner, D., Walker, T.G., Mackinnon, G.A. (Eds.), Reading Research, vol. 5. Academic Press, Orlando, FL.

Bower, G.H., Black, J.B., Turner, T.J., 1979. Scripts in memory for text. Cognitive Psychology 11, 177–220.

Britton, B.K., Graesser, A.C., 1996. Models of Understanding Text. Lawrence Erlbaum Associates, Hillsdale, NJ.

Byrd, M., 1988. Adult age differences in the ability to comprehend ambiguous sentences. Language and Communication 8, 135–146.

Carey, P., Mehler, J., Bever, T., 1970. Judging the veracity of ambiguous sentences. Journal of Verbal Learning and Verbal Behaviour 9, 243–254.

Clark, H.H., 1975. Theoretical issues in natural language processing. In: Schank. R., Nash-Webber, B. (Eds.), Proceedings of a Conference at the Massachusetts Institute of Technology. Harcourt Brace, Cambridge, MA.

Fodor, J.A., Garrett, M.F., 1967. Some syntactic determinants of sentential complexity. Perception and Psychophysics 2, 289–296.

Forster, K.I., 1981. Priming and the effects of sentence and lexical contexts on naming time: evidence for autonomous lexical processing. Quarterly Journal of Experimental Psychology 33A, 465–495.

Frith, U., 1989. Autism: Explaining the Enigma. Blackwell, Oxford.

Frith, U., Snowling, M., 1983. Reading for meaning and reading for sound in autistic and dyslexic children. British Journal of Developmental Psychology 1, 329–342.

Gernsbacher, M.A., Goldsmith, H.H., Robertson, R.R.W., 1992. Do readers mentally represent characters' emotional states? Cognition and Emotion 6, 89–111.

Glenberg, A.M., Meyer, M., Lindem, K., 1987. Mental models contribute to foregrounding during text comprehension. Journal of Memory and Language 26, 69–83.

Halldorson et al., 1992. Author please supply details.

Happé, F.G.E., 1993. Communicative competence and theory of mind in autism: a test of relevance theory. Cognition 48, 101–119.

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.

Haviland, S.E., Clark, H.H., 1974. What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behaviour 13, 512–521.

Hoppe, R.A., Kess, J.F., 1986. Biasing thematic contexts for ambiguous sentences in a dichotic listening task. Experimental Journal of Psycholinguistic Research 15, 225–241.

Jolliffe, T., Baron-Cohen, S., 1997. Are people with autism or asperger syndrome faster than normal on the embedded figures test? Journal of Child Psychology and Psychiatry 38, S27–S34.

Jolliffe, T., Lansdown, R., Robinson, C., 1992. Autism: a personal account. Communication 26, 12–19.

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.

Kreunz, R.J., MacNealy, M.S., 1996. Empirical Approaches to Literature and Aesthetics. Advances in Discourse Processes. Erlbaum, Mahwah, NJ.

Lockyer, L., Rutter, M., 1969. A five to fifteen-year follow-up study of infantile psychosis. III. Psychological aspects. British Journal of Psychiatry 115, 865–882.

Mandler, J.M., 1984. Stories, Scripts, and Scenes: Aspects of Schema Theory. Erlbaum, Hillsdale, NJ.

McKoon, G., Ratcliff, R., 1992. Inference during reading. Psychological Review 99, 440–466.

Minshew, N.J., Goldstein, G., Siegel, D.J., 1995. Speech and language in high-functioning autistic individuals. Neuropsychology 9, 255–261.

Morrow, D., Bower, G., Greenspan, S., 1989. Updating situation models during narrative comprehension. Journal of Memory and Language 28, 292–312.

O'Brien, E.J., 1995. Automatic Components of Discourse Comprehension. Erlbaum, Hillsdale, NJ.

Onifer, W., Swinney, D.A., 1981. Accessing lexical ambiguities during sentences comprehension: effects of frequency of meaning and contextual bias. Memory and Cognition 9, 225–236.

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.

Prior, M.R., Hall, L.C., 1979. Comprehension of transitive and intransitive phrases by autistic, retarded and normal children. Journal of Communication Disorders 12, 103–111.

Prutting, C., 1982. Pragmatics as social competence. Journal of Speech and Hearing Disorders 47, 123–134.

Raven, J., Court, J., Raven, J., 1992. The Standard Progressive Matrices. Oxford Psychologists Press, Oxford.

Russell, J., 1997. Autism as an Executive Disorder. Oxford University Press, Oxford.

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.

Rutter, M., Bartak, L., 1973. Special educational treatment of autistic children: a comparative study. II. Follow-up findings and implications for services. Journal of Child Psychology and Psychiatry 14, 241–270.

Schank, R.C., Abelson, R.P., 1977. Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures. Erlbaum, Hillsdale, NJ.

Scott, F., Baron-Cohen, S., 1996. Logical, analogical, and psychological reasoning in autism: a test of Cosmides theory. Development and Psychopathology 8, 400–411.

Scott, F., Baron-Cohen, S., Leslie, A., 1999. 'If pigs could fly': an examination of imagination and counterfactual reasoning in autism. British Journal of Developmental Psychology, in press.

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.

Simpson, G.B., 1984. Lexical ambiguity and its role in models of word recognition. Psychological Bulletin 96, 316–340.

Singer, M., 1993. Causal bridging inferences: validating consistent and inconsistent sequences. Canadian Journal of Experimental Psychology 47, 340–359.

Singer, M., 1995. Causal Validation and Causal Comprehension. Erlbaum, Hillsdale, NJ.

Singer, M., Revlin, R., Halldorson, M., 1990. Bridging-inferences and enthymeme. In: Graesser, A., Bower, G. (Eds.), The Psychology of Learning and Motivation, Vol. 25. Academic Press, New York, pp. 35–51.

Snowling, M., Frith, U., 1986. Comprehension in 'Hyperlexic' readers. Journal of Experimental Child Psychology 42, 392–415.

Stanovich, K.E., West, R.F., 1981. The effect of sentence context on ongoing word recognition: tests of a two-process theory. Journal of Experimental Psychology: Human Perception and Performance 7, 658–672.

Stanovich, K.E., West, R.F., 1983. On priming by a sentence context. Journal of Experimental Psychology: General 112, 1–36.

Surian, L., Baron-Cohen, S., Van der Lely, H., 1996. Are children with autism deaf to Gricean Maxims. Cognitive Neuropsychiatry 1, 55–72.

Tager-Flusberg, H., 1981a. On the nature of linguistic functioning in early infantile autism. Journal of Autism and Developmental Disorders 11, 45–56.

Tager-Flusberg, H., 1981b. Sentence comprehension in autistic children. Applied Psycholinguistics 2, 5–24.

Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M., Chadwick-Dias, A., 1990. A longitudinal study of language acquisition in autistic and Down syndrome children. Journal of Autism and Developmental Disorders 20, 1–21.

Tantam, D., 1991. Asperger syndrome in adulthood. In: Frith, U. (Ed.), Autism and Asperger Syndrome. Cambridge University Press, Cambridge, pp. 147–183.

Terman, L.M., Merrill, M.A., 1973. Stanford-Binet Intelligence Scale Form L–M, 3rd edn. Houghton Mifflin, Boston, MA.

Thorndike, P.W., 1976. The role of inferences in discourse comprehension. Journal of Verbal Learning and Verbal Behaviour 15, 437–446.

Thorndike, R.P., Hagen, E.L., Sattler, J.M., 1986. The Stanford-Binet Intelligence Scale, 4th edn. Riverside Publishing Company, Chicago, IL.

Van den Broek, P., Risden, K., Husebye-Hartmann, E., 1995. The role of readers' standards for coherence in the generation of inferences during reading. In: Lorch, R.F., O'Brien, E.J. (Eds.), Sources of Coherence in Reading. Erlbaum, Hillsdale, NJ, pp. 353–373.

Wechsler, D., 1981. Wechsler Adult Intelligence Scale – Revised. The Psychological Corporation, New York.

Wing, L., 1981. Asperger Syndrome: a clinical account. Psychological Medicine 11, 115–129.

World Health Organisation, 1992. International classification of diseases and related health problems (ICD-10), 10th edn. World Health Organisation, Geneva.