British Journal of Developmental Psychology (2005), 23, 383–399

The British Psychological Society

www.bpsjournals.co.uk

The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder

Courtenay Frazier Norbury*

Oxford Study of Children's Communication Impairments, University of Oxford, UK

Happé (1993) proposed that theory of mind (ToM) understanding was necessary for comprehension of metaphorical expressions. The current study investigated the role of both ToM and language ability in metaphor understanding. Ninety-four children aged 8–15 years with communication impairments were grouped according to language ability and autistic symptomatology in the first instance, and then according to ToM performance. Their performance on a metaphor task was compared to 34 typically developing age-matched peers. These analyses showed that only children with language impairment, with or without concurrent autistic features, were impaired on the metaphor task. Furthermore, possession of first-order ToM skills did not ensure metaphor comprehension. Instead, semantic ability was a stronger predictor of performance on the metaphor task. These results are considered with reference to the view that ToM understanding is necessary for the comprehension of metaphor.

The diagnosis of autism is dependent in part on abnormalities of communication. Children with core autism exhibit delayed and/or impaired language development, whereas children with Asperger's syndrome reach developmental language milestones on time. However, both groups experience persistent abnormalities of pragmatic language, or the use of language in meaningful contexts. These pragmatic difficulties are evident in conversation (Adams, Green, Gilchrist, & Cox, 2002), understanding jokes (Ozonoff & Miller, 1996), figurative expressions (Dennis, Lazenby, & Lockyer, 2001; Minshew, Goldstein, & Siegal, 1995), and inferencing (Jolliffe & Baron-Cohen, 1999, 2000).

These pragmatic difficulties are often attributed to deficits in theory of mind (ToM) understanding, defined as the ability to appreciate the mental states of others (see Baron-Cohen, Tager-Flusberg, & Cohen, 2000 for a review). Happé (1993) used Relevance Theory (Sperber & Wilson, 1995) to make explicit connections between pragmatic communication and ToM ability. Relevance theory highlights the importance of understanding speaker intention in communication, as a means to make sense of often

* *Correspondence should be addressed to Courtenay Frazier Norbury, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK (e-mail: courtenay.norbury@psy.ox.ac.uk).*

DOI:10.1348/026151005X26732

ambiguous language. On this view, autistic individuals should be most impaired on pragmatic language tasks in which attributing intentions is necessary for comprehension, because of their underlying deficits in theory of mind (Happé, 1993; Leinonen & Kerbel, 1999). Happé argued that comparing similes and metaphors provided a good test case for this theory because although they could be conceptually identical, similes and metaphors vary in the degree to which speaker intention must be evoked in order to interpret the expression correctly.

Specifically, she argued that one did not need to access speaker intention to understand similes ('The room was *like* an oven'), because these expressions are literally true. Metaphors ('The room was an oven') on the other hand, are not literally true, and would therefore require the listener to infer the speaker's intention that the statement be interpreted in a non-literal fashion. Happé predicted that first order ToM was necessary to understand metaphor, but similes could be processed by individuals who were not able to pass first order ToM tasks.

She tested this prediction using a sentence completion task that included synonyms, similes, and metaphor. The similes and metaphors were conceptually identical and differed only in that the similes had the test phrase 'he was like. . .', whereas the metaphor phrase was 'he really was. . .'. Eighteen individuals with autism (aged 9–28 years) were divided on the basis of their ToM scores into no-ToM, first-order and second-order groups. The no-ToM group did not differ from the other two groups on the synonym and simile conditions, but did achieve significantly lower scores on the metaphor condition. The first- and second-order groups did not differ from one another. Furthermore, the no-ToM group found metaphor more difficult than similes, whereas there was no such difference within the other two groups. These results were interpreted as evidence for 'the necessity of representing intention for understanding metaphor. . .' (p. 109).

Theory of mind and metaphor: Necessary but not sufficient?

Although Happé made a compelling argument for the role of ToM in metaphor comprehension, it clearly cannot be the whole story. Children typically pass first-order false belief tasks at around the age of 4 and yet understanding of metaphor increases throughout adolescence and young adulthood (Nippold, 1998). Children with language impairments have deficits in understanding metaphorical language (Highnam, Wegmann, & Woods, 1999; Rinaldi, 2000), even though they are reported to have relatively intact ToM abilities (Leslie & Frith, 1988; Shields, Varley, Broks, & Simpson, 1996; Ziatas, Durkin, & Pratt, 1998; though see Farmer, 2000; Miller, 2001 for evidence of ToM delay in children with specific language impairment – SLI). Both typically developing children and children with language impairments therefore provide evidence that ToM is not sufficient to ensure metaphor understanding.

Language ability and metaphor comprehension: A developmental perspective

Similes and metaphors both require individuals to see similarities between two entities that are normally considered distinct. This is a complex developmental process that requires the acquisition of a number of skills. In the first place, children must acquire sufficient world knowledge and have broad enough semantic representations to capture the comparison being made. Often the relevant features for comparison are not the most salient features of either entity. To illustrate this point, consider the metaphor, 'Lawyers

are sharks.' In order to interpret this metaphor successfully, one must know something about lawyers, and at the same time know more about sharks than that they are animals, they swim, have fins, and so forth. This additional information is likely to be more abstract, that is, they are ferocious and aggressive. Then one must decide in what respect sharks and lawyers are similar, which may involve analogical reasoning skills. Gernsbacher and Robertson (1999) also suggest that individuals must enhance the relevant features (ferocious, aggressive) while at the same time suppressing irrelevant knowledge (fins, sharp teeth). Given this example, it is easy to see that metaphor comprehension is not an all or nothing phenomenon and that success will depend in part on the content of a particular metaphor. As Vosniadou (1987) points out, '. . . it is quite unlikely that a metaphor will be understood if it compares concepts children or adults know little about' (p. 870).

Evans and Gamble (1988) provide further evidence for the role of semantic knowledge in metaphor comprehension. They compiled a list of all the entities in their metaphor task and asked children to list the salient features of each item. Six weeks later, they asked children to define metaphors containing those entities. They found that children's comprehension errors occurred when they had provided incomplete or irrelevant information in the salient features list.

Gentner (1988) compared different types of metaphorical understanding across wide age ranges.[1] Attributive metaphors ('a cloud is a marshmallow') were those in which the comparison between two entities was based on shared physical properties. Relational metaphors ('the moon is a lightbulb') involved more abstract knowledge about non-physical characteristics such as function. Double metaphors allowed either interpretation. For instance, 'plant stems are drinking straws' could be analysed either by shared physical characteristics (long, thin) or relational characteristics (both draw up water). Gentner asked children and adults to interpret each type of metaphor and found that younger children focused more on attributional properties in defining all types of metaphor. In contrast, adults preferred relational interpretations for double metaphors. This suggests that as children acquire more abstract semantic knowledge, they are better able to support metaphor comprehension.

A consistent finding in the developmental literature is that young children find similes easier to understand than metaphors, even if they are conceptually identical (Reynolds & Ortony, 1980; Vosniadou, 1987; Vosniadou, Ortony, Reynolds, & Wilson, 1984). The children involved are typically at an age when they should pass first-order ToM tasks. This is at odds with the finding in the autism literature that those individuals passing first-order ToM tasks understood metaphor as well as they understood simile. Reynolds and Ortony (1980) suggested that similes may be easier to understand because they contain an explicit syntactic clue that a comparison should be made. Early on, children may interpret metaphors literally, but this tendency decreases with age. Furthermore, these studies have shown that context boosts children's interpretation of metaphorical statements, partly because a literal interpretation would be inappropriate and nonsensical in such a context (Vosniadou, 1987).

In summary, the literature reviewed here highlights the importance of semantic and conceptual knowledge in metaphor comprehension. It is not surprising that children with language impairments find metaphor challenging, because they frequently have poor vocabularies and impoverished semantic representations for the words they do

[1] *She presented all items as similes, although she called them metaphors.*

have. The semantic skills of individuals with autism have rarely been assessed in any detail, though there is some evidence that despite basic word knowledge (i.e. category representations and associations) individuals with autism do not apply this knowledge spontaneously (Tager-Flusberg, 1985, 1991; Bowler, Gardiner, Grice, & Saavalainen, 2000). Therefore, it would seem possible that semantic ability could contribute to individual differences in metaphor comprehension in children with autistic disorder. Indeed, those individuals in Happé's study who were the most successful on the metaphor task were also the most verbally able. This suggests that at least some children with autism may have semantic deficits and that their poor performance with metaphors may be due to problems with language rather than ToM. One problem for interpreting these data is that verbal ability and performance on ToM tasks are highly correlated (Happé, 1995), making it difficult to tease apart the relative contributions each one makes to metaphor competence.

Predictions of the current study

The current study explored semantic knowledge, theory of mind ability, and metaphor comprehension across groups of children with developmental language disorder, autistic spectrum disorder only, or a combination of the two. The following hypotheses were contrasted:

(1) In keeping with Happé (1993), the first hypothesis is that ToM ability is an important determinant of metaphor comprehension. Therefore, children who pass first- and second-order ToM tasks should have a better understanding of metaphor than those who do not, regardless of language status. In addition, ToM ability should predict unique variance in metaphor comprehension once language ability has been taken into account.
(2) The second hypothesis asserts a primary role for language competence, and in particular semantic knowledge, in metaphor comprehension. On this view, children with language impairment should have more difficulty understanding similes and metaphors than those children who do not have language difficulties, regardless of autistic status or ToM ability. Furthermore, ToM ability should not predict unique variance in metaphor comprehension once semantic knowledge has been taken into account.

Methods

Participants

Ninety-four children with communication impairments were recruited from specialist schools and units throughout the southeast of England. Children were selected who met the following criteria: aged between 8 and 15 years, non-verbal abilities within the normal range standard score of 80 or above on the performance subscale of the Wechsler abbreviated scales of intelligence (Wechsler, 1999), monolingual English-speaking home environment, no sensorineural hearing impairment, and no evidence of neurological impairment. Thirty-four typically developing children matched for age and non-verbal ability were also selected.

Children were selected to represent a range of clinical diagnoses based on clinical report and documented medical diagnoses available in school files. These included children with SLI ($N = 19$), children with primary pragmatic language impairment

(PLI, $N = 15$), children with PLI and autistic behaviours (more consistent with a diagnosis of pervasive developmental disorder not otherwise specified, $N = 24$), children with high-functioning autism ($N = 18$) and children with Asperger's disorder ($N = 18$). Grouping children according to these diagnoses proved untenable for the following reasons. First, there is evidence to suggest that PLI (formerly semantic-pragmatic disorder or semantic-pragmatic deficit syndrome) is a variable correlate of SLI or autistic disorder, rather than a diagnostic entity itself (Bishop & Norbury, 2002; Rapin, 1996). Second, it was not possible to independently verify autistic spectrum diagnoses. This was particularly problematic in distinguishing between high-functioning autism and Asperger's disorder. Many children diagnosed with Asperger's disorder had a clear history of early language delay, which is inconsistent with current DSM-IV guidelines (American Psychiatric Association, 2000). Finally, these diagnostic labels represented extremely heterogeneous groups in terms of language ability and/or severity of autistic symptomatology. Because these two factors were of particular interest in explaining individual differences in metaphor understanding, it was decided to regroup children according to autistic status and language ability in order to elucidate rather than confound results.

Group selection procedure

Autistic status

The first question asked was, 'does the child present with autistic-type behaviours?' According to clinical records, 60 children did have reported autistic features, while 34 did not. This was independently assessed using two checklists, the Autism Screening Questionnaire (ASQ – Berument, Rutter, Lord, Pickles, & Bailey, 1999) and the Children's Communication Checklist-2 (CCC-2 – Bishop, 2003). The ASQ is a 40-item parental questionnaire that assesses all three aspects of the autistic triad: communication, social interaction, and restricted interests and behaviours. It focuses on the child's behaviour between the ages of 4 and 5 and provides good diagnostic agreement with the more extensive Autism Diagnostic Interview-Revised (ADI-*R* – Lord, Rutter, & Le Couteur, 1994; Berument *et al.*, 1999; Bishop & Norbury, 2002). The CCC-2 is designed to measure pragmatic abilities but includes two subscales (social interaction and specific interests) tapping autistic behaviour. Although the number of items is small (14), it provides a measure of current behaviour, rather than historical report (as in the ASQ).

Language status

The second question asked was, 'does the child have a structural language impairment, as measured by standardized tests?' Three tests of structural language were administered: the British picture vocabulary scales (BPVS – Dunn, Dunn, & Whetton, 1997), which measures receptive vocabulary; the Concepts and Directions subtest of the Clinical Evaluation of Language Fundamentals, UK 3rd Edition (CELF-III^{UK}, Semel, Wiig, & Secord, 2000), which measures understanding of increasingly complex sentences; and the Recalling sentences subtest of the CELF-III^{UK}, which indexes expressive language ability. Children scoring below -1.25 *SD* on two of the three tests, or -2 *SD* on one test were regarded as having a language impairment. Using this analysis, children generally fell into one of three groups: language impaired (LI, $N = 28$), autistic spectrum plus language impairment (ASL, $N = 31$) and autistic spectrum only

(ASO, $N = 29$). A small number of children ($N = 6$) did not fit into any of these groups. They appeared to have pragmatic difficulties without additional language impairment or autistic symptomatology. These children were not included in the initial group analysis, but their scores were entered in correlational analyses.

Table 1 reports the mean age, non-verbal ability scores, and scores on the selection measures for the three clinical groups and the typically developing control group. There was a marginal difference in age, $F(3, 118) = 2.34$, $p = .08$, with a trend for the control group to be younger than the ASL group. Although all children were selected to have non-verbal abilities within the normal range, the two groups with language impairment (ASL and LI) had significant lower scores than the ASO and control groups, who did not differ from one another, $F(3, 118) = 9.97$, $p < .001$. The results of the ASQ and the CCC-2 composite provide validation of clinician report in that the two groups with reported autistic features (ASL and ASO) have significantly higher scores on the ASQ, and significantly lower scores on the CCC-2 than the control and LI groups, ASQ: $F(3, 102) = 64.19$, $p < .001$; CCC-2: $F(3, 96) = 43.87$, $p < .001$. The LI group scores were also significantly different from controls ($p < .001$). It is not unusual for children with language impairment to exhibit raised thresholds of autistic type behaviour relative to typically developing children, even if they do not meet full criteria for autistic disorder (Bishop & Norbury, 2002).

As expected, there were significant group differences on all language measures, BPVS: $F(3, 118) = 53.45$, $p < .001$; Concepts: $F(3, 118) = 79.93$, $p < .001$; Recalling sentences: $F(3, 118) = 63.87$, $p < .001$. In all instances, the LI and ASL groups did not

Table 1. Mean (*SD*) of chronological age, non-verbal ability (performance subscale of Wechsler abbreviated scales of intelligence) and two checklists assessing autistic behaviours, the Autism Screening Questionnaire and the Children's Communication Checklist-2 (scales I and J combined)

	Language impaired (LI) $N = 28$	Autistic spectrum + LI (ASL) $N = 31$	Autistic spectrum only (ASO) $N = 29$	Control $N = 34$
Age (years)	11.90 (2.19)	12.04 (2.31)	11.76 (1.85)	10.83 (1.84)
WASI	97.25[a] (9.42)	99.39[a] (11.66)	108.31[b] (12.48)	110.56[b] (11.91)
Autistic status[a]				
ASQ	13.08[a] (7.31)	22.28[b] (7.55)	24.75[b] (6.12)	2.79[c] (2.36)
CCC-2 (I & J)	11.23[a] (6.69)	4.60[b] (6.01)	5.73[b] (6.00)	23.72[c] (4.96)
Language status				
BPVS	81.61[a] (10.81)	77.52[a] (11.78)	102.66[b] (13.18)	110.18[b] (12.62)
Recall	4.14[a] (1.74)	4.19[a] (1.58)	8.10[b] (2.83)	10.65[c] (2.52)
Concepts	4.68[a] (2.00)	4.16[a] (1.32)	9.00[b] (3.17)	11.91[c] (2.43)

[a] The ASQ was unavailable for 3 children in the LI group, 2 in ASL, 1 in ASO and 10 in the control group. The CCC-2 was unavailable for 2 children in the LI group, 1 in ASL, 3 in ASO and 16 in the control group.

Note. Values with different subscripts in each row are statistically different at $p < .01$. For the ASQ, higher numbers indicate increased severity of autistic symptoms. For the CCC-2, lower numbers are indicative of more severe impairment. The normative mean on this measure is 20. Mean score (*SD*) on Recall and Concepts is 10 (3). BPVS, British picture vocabulary scales, standard scores reported with a normative mean of 100 and a standard deviation of 15; Recall, Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals (CELF-III^{UK}); Concepts, Concepts and Directions subtest of the CELF-III^{UK}

differ from one another and had significantly poorer scores than both the ASO and control groups. On BPVS, the ASO and control groups did not differ from one another. However, on both the Concepts and Recalling sentences subtests, the ASO group had significantly lower scores than the control group ($p < .01$), even though their mean scores were within the normal range.

Experimental procedure

All children were assessed in a quiet room in their school or at home. Three tasks were administered: a test of semantic knowledge, a ToM task, and a metaphor comprehension task.

Semantic knowledge

The Test of Word Knowledge (ToWK – Wiig & Secord, 1992) is a standardized test measuring broader aspects of word knowledge than typical vocabulary tests such as the BPVS. There are two receptive subtests: Synonyms, in which the child must identify a word closest in meaning to a target word from a selection of four, and Figurative expressions, in which the child must select the most appropriate interpretation of an idiomatic phrase from a choice of four. In addition, there are two expressive subtests: Definitions, in which the child provides a definition of a stated word, and Multiple Contexts, in which the child must state two possible meanings of an ambiguous word such as *bat*. Scores from these four subtests are transformed into a total score with a normative mean of 100 and *SD* of 15.

Theory of mind

The ToM task comprised two stories from Sullivan, Zaitchik, and Tager-Flusberg (1994) and followed the cartoon procedure used by Farmer (2000). These stories were designed for use with preschool children and were therefore thought to be ideal for use with language impaired children because (a) they were relatively short, (b) questions were interspersed within the story to reduce memory load, (c) a memory aid was included before the crucial test question to further reduce the memory confound, and (d) a linguistic control question was included that had the same syntactic structure as the second order questions.

The first story was a variation on the ice cream van story. The two main characters, John and Mary meet the ice cream man in the park. When Mary goes to get money to buy an ice cream, John learns that the ice cream man is leaving the park to go to the school. On her return to the park, Mary meets the ice cream man again and follows him to the school. John is unaware of this meeting and therefore mistakenly believes that Mary will return to the park to buy an ice cream. The second story involves deception. A mother hides a birthday surprise (a new puppy) from her son Peter and deceives him by telling him that he will receive a new toy for his birthday instead of a puppy. However, Peter accidentally discovers the puppy downstairs. His mother is unaware of this discovery and mistakenly assumes that he believes he will not get a puppy for his birthday (see Sullivan *et al.* for details).

Each story contained two questions tapping first order knowledge ('John thinks X') and two tapping second order knowledge ('John thinks that Mary thinks X'). These were scored as correct or incorrect, for a total possible score of eight. A justification question was asked in each story after the last second-order question. Responses were scored as

correct if they made reference to second-order reasoning ('John doesn't know that Mary saw the ice cream man'), communication ('the ice-cream man told Mary he would be in the park'), or location/deception element of the story ('the ice cream man was in the park when John and Mary were there'/'mum wanted to surprise Peter with the new puppy'; Sullivan *et al.*, 1994).

Metaphor task

The metaphor task was adapted from that of Happé (1993), which compared conceptually similar metaphors and similes. Children were asked to complete six sentences in each of three conditions: synonym, simile, and metaphor. To control for possible differences in conceptual complexity of individual items, three versions of each sentence were constructed and counterbalanced across subjects. The task differed from Happé's task in several ways. First, conditions were mixed and items randomly ordered so that each participant answered all 18 items in one sitting. This was done to avoid a response bias that might result from presenting the same type of item in blocks (Reynolds & Ortony, 1980). Second, vocabulary was simplified. Finally, each item had its own choice of four words that could complete the sentence. In Happé's study, there were five sentences per condition and a total of six possible responses. It is therefore possible that able participants used a process of elimination to solve the task. Foils (the three other word choices presented for each sentence) in the current study were designed to avoid ceiling effects and included words that were related to the sentence or target (Reynolds & Ortony, 1980). An example is provided in Appendix A, and the complete set of items is listed in Appendix B.

The examiner read each sentence and all four choices aloud to the child as many times as necessary. The child was asked to circle the word that completed the sentence and helped the sentence make sense. A practice item was provided and feedback given on that item. Items were scored as correct or incorrect.

Results

Scores on the experimental measures were analysed using ANOVA with Scheffé tests used for *post hoc* comparisons unless otherwise noted. In addition, the effect size η^2, which estimates the proportion of total variance accounted for by the independent variable, is also reported.

Semantic skills and theory of mind

Group means on the ToWK and ToM tasks are shown in Table 2. A significant group difference was seen in the ToWK and the pattern of results was similar to that seen for the core language measures, $F(3, 116) = 66.77$, $p < .001$, $\eta^2 = .634$, with the ASL and LI groups achieving lower scores than the ASO and control groups (both ps $< .001$). The ASO group again had scores within normal limits, but significantly poorer than the control group ($p < .001$).

Scores on the ToM task were analysed using the non-parametric Kruskal-Wallis test, because the control group was at ceiling on this measure and there were unequal variances across groups. There was a significant group effect, $\chi^2 = 38.94$, $N = 122$, $df = 3$, $p < .001$. *Post hoc* comparisons revealed that the control group had significantly higher scores than both the LI and ASL groups ($p < .05$), but did not

Table 2. Mean (*SD*) scores per group on the Test of Word Knowledge (normative mean 100, *SD* 15) and the theory of mind task (maximum raw score 8, maximum justification raw score 2)

	Language impaired (LI) $N = 28$	Autistic Spectrum + LI (ASL) $N = 31$	Autistic Spectrum Only (ASO) $N = 29$	Control $N = 32$[a]
ToWK	71.82[a] (10.27)	71.23[a] (11.63)	92.62[b] (12.87)	110.09[c] (15.10)
ToM task	6.46[a] (1.90)	5.29[a] (3.11)	7.69[b] (0.97)	7.91[b] (0.38)
ToM justification	1.00[a] (0.80)	0.87[a] (0.82)	1.66[b] (0.61)	1.65[b] (0.54)

[a] Due to time constraints, ToWK scores were unavailable for two children in the control group.
Note. Different subscripts in each row differ significantly at $p < .001$.

differ from the ASO group. The ASO group also had significantly higher scores than the LI and ASL groups. The LI and ASL groups did not differ from one another. On the justification questions, the LI and ASL groups did not differ, but had significantly poorer scores than the ASO and control groups, who did not differ, $\chi^2 = 25.65$, $N = 122$, $df = 3$, $p < .001$.

The three clinical groups were reorganized according to ToM ability. Children in the second-order group answered all of the first- and second-order questions correctly. Children in the first-order group answered all of the first-order questions correctly, but had variable scores on the second-order questions. Children in the no-ToM group did not successfully answer all of the first-order questions. The percentages of children in each ToM group are shown in Table 3.

Table 3. Number (percentage) of children in each clinical group achieving different levels of ToM understanding

	no-ToM	1st-ToM	2nd-ToM
Language impaired	3 (10.7%)	13 (46.4%)	12 (42.9%)
Autistic spectrum + LI	9 (29.0%)	11 (35.5%)	11 (35.5%)
Autistic spectrum only	1 (3.4%)	4 (13.8%)	24 (82.8%)

Note. Of the 6 children not placed in these clinical groups, 2 were 1st-ToM and 4 were 2nd-ToM.

Table 4 shows the scores on the selection measures by ToM group. Across all tasks, a similar pattern of results is seen. The no-ToM and first-order groups do not differ and have poorer scores than both the second-order and control groups. The second-order group also has significantly poorer scores than the control group. Although their group means on semantic tasks (BPVS and ToWK) were within normal limits, they showed mild impairments on tests of sentence processing (concepts and recalling sentences).

Metaphor task

Figure 1 illustrates the outcome of the metaphor task. It can be seen that both the ASL and LI groups have slightly lower scores than the other groups on the synonym items, confirmed by a univariate ANOVA, $F(3, 118) = 5.92$, $p = .001$, $\eta^2 = .133$. *Post hoc* comparisons revealed that these groups had significantly lower scores than the control group, reflecting their poorer performance on standardized measures of semantic ability.

Table 4. Mean (*SD*) of chronological age, non-verbal ability and language performance of the ToM groups

	No-ToM $N = 13$	1st-ToM $N = 30$	2nd-ToM $N = 51$	Control $N = 34$
Age	11.00 (2.38)	11.82 (1.88)	12.09 (2.18)	10.83 (1.84)
WASI	97.85^{a} (8.60)	100.07^{a} (10.59)	103.78^{ab} (13.26)	110.56^{B} (11.91)
BPVS	76.92^{a} (12.58)	82.50^{a} (13.21)	94.31^{b} (16.11)	110.18^{c} (12.62)
ToWK	66.69^{a} (9.82)	72.53^{a} (11.57)	86.76^{b} (14.75)	110.09^{c} (15.10)
Recall	3.77^{a} (1.36)	5.13^{a} (2.13)	6.63^{b} (3.10)	10.65^{c} (2.52)
Concepts	3.23^{a} (0.83)	5.53^{a} (2.60)	7.02^{b} (3.48)	11.91^{c} (2.43)

Note. Items with differing subscripts in the same row are significantly different at $p < .01$.
WASI, BPVS, and ToWK have a normative mean of 100 and *SD* of 15. Recall and Concepts have normative mean of 10, *SD* of 3.

Differences between simile and metaphor understanding were analysed using a repeated measures ANOVA, with group as the between subjects factor and question type (simile vs. metaphor) as the within subject factor. This analysis indicated a main effect of question type, $F(1, 118) = 72.01$, $p < .001$, $\eta^2 = .379$, and a main effect of group, $F(3, 118) = 8.46$, $p < .001$, $\eta^2 = .177$. The Group × Question type interaction was not significant ($F = 1.21$). Figure 1 shows that all groups find metaphors more difficult than similes. *Post hoc* analysis of the group effect revealed that the control group outscored both the LI and ASL groups on similes, and the control and ASO groups had significantly better scores on metaphor items than the ASL group.

Reclassification of groups according to ToM performance: Metaphor understanding

In order to compare these findings with Happé (1993), the analysis was rerun with the groups reorganized according to ToM performance, as shown in Fig. 2. Here, the no-ToM group had significantly poorer synonym scores than all other groups. The repeated measures ANOVA again showed a main effect of question type, $F(1, 118) = 73.18$, $p < .001$, $\eta^2 = .383$, and a main effect of group, $F(3, 118) = 12.61$, $p < .001$, $\eta^2 = .243$. The Group × Question type interaction was non-significant, $F(3, 118) = 1.95$, $p = .13$. All groups found metaphor more difficult than simile.

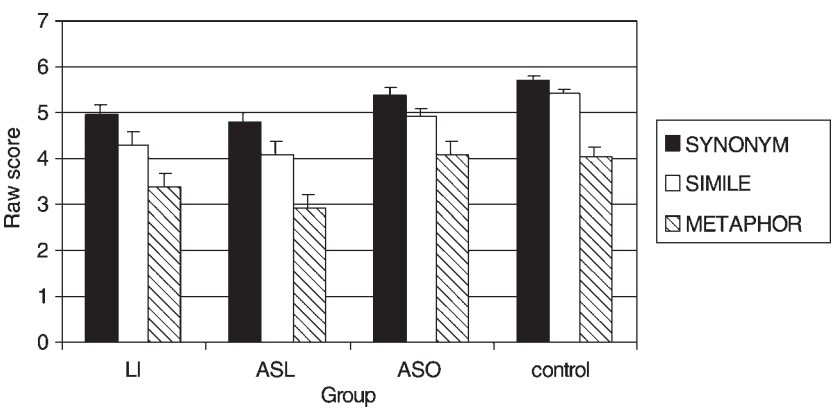

Figure 1. Performance on the metaphor task by clinical group. Mean raw scores (+ SE) on synonym, simile and metaphor items are depicted.

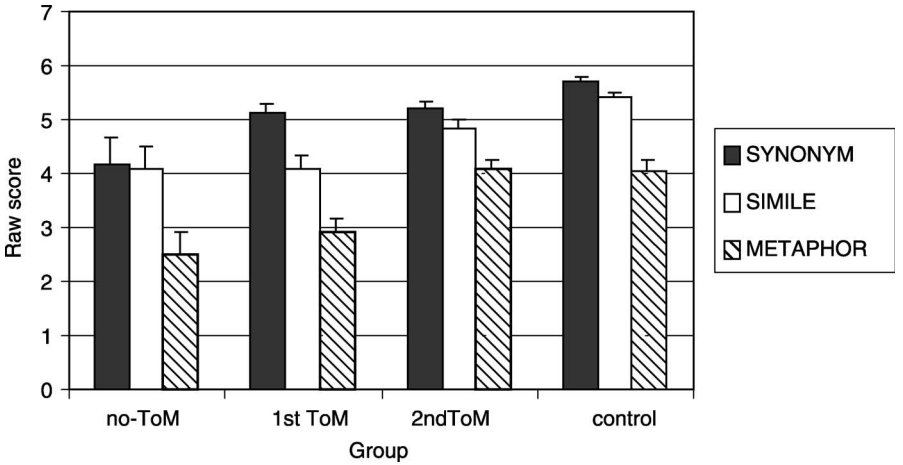

Figure 2. Performance on metaphor task by ToM group. Mean raw scores (+ SE) for synonym, simile, and metaphor items are depicted.

Contrary to Happé's findings, the no-ToM and first-order ToM groups did not differ on metaphor items, and were significantly more impaired than the second-order group, who did not differ from controls. The group differences remained when age and non-verbal ability were covaried.

Relationships between metaphor comprehension, language and ToM measures

Table 5 reports the correlation coefficients between metaphor comprehension, ToM scores, ToWK scores, and selection measures. Because the control group was at ceiling on the metaphor and ToM task, they were not included in this analysis. It is apparent that both language and ToM abilities correlate significantly with metaphor comprehension, whereas autistic symptomatology (as measured by the ASQ) does not. However, ToM ability itself correlates significantly with all language measures, making it difficult to identify the independent contributions language and ToM make to metaphor understanding. A hierarchical regression analysis was conducted to address this issue.

A typical research strategy in the autism literature is to match participants on verbal mental age, using a receptive vocabulary test such as the BPVS. One of the hypotheses

Table 5. Correlation coefficients ($df = 94$) between background measures, metaphor comprehension and theory of mind scores within the clinical groups

	Metaphor comprehension	Theory of mind score
Age	.29**	.22*
Non-verbal ability	.17	.18
BPVS	.27**	.38***
ToWK	.41***	.47***
Concepts	.40***	.37***
Recall	.26*	.40***
ASQ^{+}	−.06	.03
Theory of mind	.39***	

Note. * Indicates significant value at $p < .05$, ** $p < .01$, ***$p < .001$, $+df = 88$.

for the current study was that broader and more flexible semantic knowledge was important for understanding metaphor. Therefore, it was necessary to establish what additional contribution semantic knowledge made to metaphor comprehension once basic vocabulary knowledge was taken into account. This resulted in the selection of the following five predictor variables ($N = 94$): age, BPVS, ToWK, ToM, and ASQ scores. The outcome variable was accuracy on metaphor items. The results are shown in Table 6. Age was entered first in the regression, accounting for a significant 7% of the variance. BPVS scores were entered at the next stage, explaining an additional 15% of the variance. The addition of ToWK scores explained a further 12% of the variance. However, examination of the β coefficients indicated that the effect of the BPVS was no longer significant once ToWK scores were included in the model. Entering ToM scores at this point did not account significantly for additional variance. The addition of the ASQ accounted for a small 3% of variance, which just failed to reach significance ($p = .06$). This final model accounted for 37% of the variance in metaphor understanding.

Table 6. Regression analyses predicting metaphor comprehension from age, language ability, theory of mind and ASQ scores

Step	Predictor	R^2	R^2 change	*F* value significance	Standardized β coefficients	*t* value	*p* significance
1	Age	.074	.074	0.01	0.428	4.17	< .001
2	BPVS	.221	.147	< 0.001	0.004	0.03	.98
3	ToWK	.337	.116	< 0.001	0.538	3.55	.001
4	Theory of mind	.341	.004	0.49	0.050	0.47	.64
5	ASQ	.369	.028	0.06	−0.173	−1.91	.06

Discussion

The study reported here was motivated by claims that first-order ToM understanding is necessary for metaphor comprehension. Two methods of analysis were employed to test this claim. First, group comparisons were made amongst groups that varied in terms of autistic severity, language impairment, and ToM understanding. Second, regression analyses were used to pinpoint the unique contributions of semantic knowledge and ToM ability to metaphor understanding. These lines of enquiry converge on the same conclusion: semantic knowledge is a key determinant of metaphor comprehension. In the first analysis, two groups of children with language impairment had more difficulty with metaphor understanding than other children with autistic spectrum disorder but no additional language difficulties. Furthermore, children passing first-order ToM tasks were no more successful at metaphor than the no-ToM group. Importantly, both of these groups had significantly poorer scores on the semantic task than the second-order ToM group, who achieved significantly higher metaphor scores. In the second analysis, broad semantic knowledge, as measured by the ToWK, predicted a significant amount of variance in metaphor comprehension, whereas ToM understanding and severity of autistic symptoms did not.

These findings are in keeping with developmental studies that found a relationship between semantic skills and metaphor understanding (Evans & Gamble, 1988; Gentner, 1988). In line with Happé (1993), individuals with autism who found metaphor most

challenging also had the weakest verbal abilities. However, Happé argued that differences in metaphor understanding could not be solely attributed to verbal ability, because a control group of individuals with intellectual impairment succeeded at both ToM and metaphor tasks, despite having verbal mental ages equivalent to the no-ToM group. One possible explanation for these findings may be the way in which verbal ability was assessed. Receptive vocabulary is a crude index of semantic knowledge and the results reported here suggest that the BPVS may overestimate semantic competence in clinical populations. Although the ASO group was matched to controls on BPVS in this study, they had significantly poorer scores on the ToWK. Furthermore, the BPVS did not predict unique variance in metaphor comprehension once ToWK scores had been taken into account, whereas ToWK continued to predict unique variance. It could be that the control group Happé used differed from the no-ToM group on other aspects of semantic ability and that this assisted them in metaphor comprehension.

The findings reported here differed from Happé's in two important respects. First, Happé's model predicted that those in the first ToM group would perform equivalently to the second ToM group and significantly better than the no-ToM group. This was not the case. Both no- and first-order ToM groups were impaired relative to those with second-order ToM skills. Second, Happé found that simile and metaphor understanding did not differ within the first- and second-order groups, only the no-ToM group found metaphor more difficult than simile. In this study all groups, including controls, found metaphors more difficult than similes.

What accounts for the differences between studies? The first difference may relate to the items and foils used. In Happé's study, there were six foils for all five sentences in each condition, one for each sentence and one extra. Participants were told they could use answers more than once, but it is likely that able participants used a process of elimination to work out successive answers. Furthermore, the nature of the task ensured that the choices were disparate, which could have facilitated metaphor selection. In this study, each item had four choices and the three foils were related to the sentence in some way. This made the task quite challenging and may explain why individuals in the first ToM group in the current study (who had language deficits equivalent to the no-ToM group) were impaired on this task. However, as both similes and metaphors had the same foils, the foils alone cannot explain the difference in performance between the two sentence types.

Another important difference between studies is that the participants were considerably older in Happé's study. It is well established that metaphor understanding improves with age, and age was a significant predictor of metaphor understanding in the current study. It is possible that the increased experience with and exposure to metaphor may have benefited the participants in Happé's study.

Why are similes easier?

The developmental literature has suggested that similes may be easier for young children to understand because similes contain an explicit syntactic cue that a comparison is necessary (Reynolds & Ortony, 1980). An examination of the errors children made to metaphors in this study suggests that in the absence of such a clue, children may search for an answer that bears a literal relationship to the sentence. For instance, in the item, 'Joe spent too long in the swimming pool' the expected metaphor response was, 'He was a prune' but many children chose, 'He was a float', a float being more literally

associated with a swimming pool than a prune. They did not tend to make the same mistakes with similes, because, 'Joe was like a prune' is perfectly reasonable.

Is theory of mind necessary for metaphor comprehension?

Two possible outcomes for this study were outlined in the introduction. The first considered the role of ToM in metaphor comprehension. Happé (1993) used Relevance Theory to argue that metaphor understanding required an appreciation of speaker intention, and that therefore one would need to pass first-order theory of mind in order to understand metaphor. The results of the current study suggest that ToM is not sufficient for metaphor understanding. However, it is not possible to argue that ToM is unnecessary for understanding metaphor, not least because ToM and verbal skills are so closely related. It would be reasonable to assume that understanding speaker intention may make the task easier in natural contexts, but the results of this study suggest it is not sufficient to ensure adequate metaphor comprehension.

The second hypothesis was that language ability in general and semantic skills specifically are important for metaphor comprehension. The results reported here provide evidence in support of this hypothesis. The use of clinical groups that varied on dimensions of autistic behaviour and language impairment highlights an important factor for future research. Only those individuals with autism who had concomitant structural language deficits were impaired on the metaphor task and they were indistinguishable from children with language impairment who did not have clinically significant autistic features. This finding suggests that at least some of the pragmatic deficits characteristic of autism may be attributable to lower level linguistic deficits. More detailed investigations of the language skills of autistic individuals participating in research studies are warranted (cf. Tager-Flusberg & Joseph, 2003) and the use of non-autistic children with language impairment as a comparison group would further elucidate the nature of pragmatic impairments in both disorders.

There are clearly a number of top-down (contextual processing, world knowledge, and experience) and bottom-up (semantic analysis) processes that work synergistically to arrive at metaphor understanding. In highly verbal individuals with autism, superior semantic knowledge may help them to circumvent any deficits in ToM understanding they may have. For language impaired individuals, improving semantic representations and making comparisons more explicit will enable these children to gain a better understanding of metaphorical language.

Acknowledgements

This work was completed whilst the author was in possession of a Wellcome Trust Prize Studentship and submitted in partial fulfilment of a Doctor of Philosophy degree at the University of Oxford. I would like to thank Dorothy Bishop and Kate Nation for their insightful comments and helpful suggestions on an earlier draft of this paper. I also thank Professor Sir Michael Rutter for making a pre-publication version of the ASQ available. A very warm thank you is extended to all of the children and families taking part in this study, and to the following schools who have assisted in recruitment of participants and hosting this research: Barley Hill Primary, Blossom House, Chinnor Resource Base, Donnington Language Resource Base, Gap House, Hope Lodge, John Horniman School, the Link Primary and Secondary schools, Moor House, More House, Oxted County School, St. Catherine's School, St. Dominic's School, Sir Thomas Abney Language Resource Base, Whitefields School.

References

Adams, C., Green, J., Gilchrist, A., & Cox, A. (2002). Conversational behaviour of children with Asperger syndrome and conduct disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *43*, 679–690.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Baron-Cohen, S., Tager-Flusberg, H. & Cohen, D. J. (Eds.), (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). Oxford: Oxford University Press.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry*, *175*, 444–451.

Bishop, D. V. M. (2003). *The children's communication checklist – 2*. London: The Psychological Corporation.

Bishop, D. V. M., & Norbury, C. F. (2002). Exploring the borderlands of autistic disorder and specific language impairment: A study using standardised diagnostic instruments. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *43*, 917–931.

Bowler, D. M., Gardiner, J. M., Grice, S., & Saavalainen, P. (2000). Memory illusions: False recall and recognition in adults with Asperger's syndrome. *Journal of Abnormal Psychology*, *109*, 663–672.

Dennis, M., Lazenby, A. L., & Lockyer, L. (2001). Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders*, *31*, 47–54.

Dunn, L. M., Dunn, L. M., & Whetton, C. W. (1997). *The British picture vocabulary scale* (2nd ed.). Windsor: NFER-Nelson.

Evans, M. A., & Gamble, D. L. (1988). Attribute saliency and metaphor interpretation in school-age children. *Journal of Child Language*, *15*, 435–449.

Farmer, M. (2000). Language and social cognition in children with specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 627–636.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *59*, 47–59.

Gernsbacher, M. A., & Robertson, R. (1999). The role of suppression in figurative language comprehension. *Journal of Pragmatics*, *31*, 1619–1630.

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. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, *66*, 843–855.

Highnam, C., Wegmann, J., & Woods, J. (1999). Visual and verbal metaphors among children with typical language and language disorders. *Journal of Communication Disorders*, *32*, 25–35.

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

Jolliffe, T., & Baron-Cohen, S. (2000). Linguistic processing in high-functioning adults with autism or Asperger's syndrome. Is global coherence impaired? *Psychological Medicine*, *30*, 1169–1187.

Leinonen, E., & Kerbel, D. (1999). Relevance theory and pragmatic impairment. *International Journal of Language and Communication Disorders*, *34*, 367–390.

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, *6*, 315–324.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659–685.

Miller, C. A. (2001). False belief understanding in children with specific language impairment. *Journal of Communication Disorders*, *34*, 73–86.

Minshew, N. J., Goldstein, G., & Siegal, D. J. (1995). Speech and language in high-functioning autistic individuals. *Neuropsychology, 9*, 255–261.

Nippold, M. A. (1998). *Later language development* (2nd ed.). Austin, Texas: PRO-ED.

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

Rapin, I. (1996). Practitioner review: Developmental language disorders: A clinical update. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37*, 643–655.

Reynolds, R. E., & Ortony, A. (1980). Some issues in the measurement of children's comprehension of metaphorical language. *Child Development, 51*, 1110–1119.

Rinaldi, W. (2000). Pragmatic comprehension in secondary school-aged students with specific developmental language disorder. *International Journal of Language and Communication Disorders, 35*, 1–29.

Semel, E. M., Wiig, E. H., & Secord, W. (2000). *Clinical evaluation of language fundamentals* (3rd UK ed.). London: The Psychological Corporation.

Shields, J., Varley, R., Broks, P., & Simpson, A. (1996). Social cognition in developmental language disorders and high-level autism. *Developmental Medicine and Child Neurology, 38*, 487–495.

Sperber, D., & Wilson, D. (1995). *Relevance* (2nd ed.). Oxford: Blackwell.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30*, 395–402.

Tager-Flusberg, H. (1985). Basic level and superordinate level categorization by autistic, mentally retarded and normal children. *Journal of Experimental Child Psychology, 40*, 450–469.

Tager-Flusberg, H. (1991). Semantic processing in the free recall of autistic children: Further evidence for a cognitive deficit. *British Journal of Developmental Psychology, 9*, 417–430.

Tager-Flusberg, H., & Joseph, R. M. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London, 358*, 303–314.

Vosniadou, S. (1987). Children and metaphors. *Child Development, 58*, 870–885.

Vosniadou, S., Ortony, A., Reynolds, R., & Wilson, P. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Child Development, 55*, 1588–1606.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. San Antonio: The Psychological Corporation.

Wiig, E. H., & Secord, W. (1992). *Test of word knowledge*. London: The Psychological Corporation.

Ziatas, K., Durkin, K., & Pratt, C. (1998). Belief term development in children with autism, Asperger syndrome, specific language impairment, and normal development: Links to theory of mind development. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*, 755–763.

Received 19 August 2003; revised version received 15 June 2004

Appendix A. Example of item and foil choices. The correct answer is in bold type.

The heating had been left on over night and the room was very warm.

(a) **synonym:**	*It was:*	**hot**	a blanket	a grill	spicy
(b) **simile:**	*It was like:*	**an oven**	a blanket	a grill	a spice
(c) **metaphor:**	*It was:*	**an oven**	a blanket	a grill	a spice

Appendix B. Sentence stimuli and synonym/metaphors used in metaphor task.

Sentence stimuli	Synonym	Metaphor
Simon had been walking in the snow for hours. His feet were	freezing	ice
The tree in my garden has grown a lot this year. It is	tall	a tower
Mum left the bread out overnight. This morning it was	hard	a brick
My school friend always protects me from bullies. He is	brave	a soldier
Laura talks so softly you can barely hear her. She is	quiet	a mouse
Jen always gets good marks on her exams. She is	clever	a professor
Father was very cross when I got home late. He was	angry	a volcano
Peter can lift very heavy weights with no problem. He is	strong	an ox
Our new school is very big and I always get lost. It is	confusing	a maze
Julie's long fingernails were painted red and gold. They were	colourful	jewels
Lou was always happy and made everyone feel good. He was	cheerful	the sun
Pat has very long and smooth hair. It is	straight	spaghetti
Joe spent too long in the swimming pool. He was	wrinkled	a prune
Sam's new pet dog is very big. It is	huge	an elephant
Louise had been shouting and crying for hours. She was	upset	a storm
The heating was on for hours and the room was warm. It was	hot	an oven
Kate had a lovely face and pretty eyes. She was	beautiful	a painting
Julian was hiding behind the tree and not moving. He was	still	a statue

Note. In the simile condition, the carrier phrase ended with the word *like*.