Infant and Child Development
Inf. Child Dev. **11**: 171–195 (2002)
DOI: 10.1002/icd.304

The Relation between Theory of Mind and Rule Use: Evidence from Persons with Autism-Spectrum Disorders

Philip David Zelazo[a,*], Sophie Jacques[a], Jacob A. Burack[b] and Douglas Frye[c]
[a] *University of Toronto, Toronto, Canada*
[b] *McGill University, Montreal, Canada*
[c] *University of Pennsylvania, USA*

Cognitive complexity and control (CCC) theory, which is a theory of executive function and its development, provides a metric for comparing task demands across domains. This metric allowed us to examine the relation between theory of mind (ToM) and one aspect of executive function, rule use, in 22 individuals with autism-spectrum disorders, including 12 severely impaired (VIQ $\leqslant$ 40; mean VMA = 4.07 years; mean CA = 17.47) and 10 mildly impaired (VIQ>40; mean VMA = 6.15 years; mean CA = 10.30) individuals. For severely impaired individuals, ToM performance was unrelated to rule use, $r = -0.40$, $p > 0.05$). However, for mildly impaired individuals, the correlation between ToM and rule use was high, $r = 0.82$, $p < 0.01$). This latter finding challenges the hypothesis of a domain-specific, ToM module, and suggests instead that poor performance on ToM tasks may be attributed to a more general difficulty using higher order rules to integrate 2 incompatible perspectives into a single system of inferences. Copyright © 2002 John Wiley & Sons, Ltd.

Key words: executive function; rule use; theory of mind; autism

A central tenet of developmental approaches to psychopathology is that the study of psychopathology can provide insights into the nature of typical developmental processes (Cicchetti, 1984; Hodapp and Burack, 1990; Burack, 1997; Luthar *et al.*, 1997). For example, psychopathologies often produce dissociations among psychological processes that typically develop together, thereby facilitating the analysis of complex functions and allowing researchers to identify those processes that continue to be related to one another even in cases of markedly atypical development.

*Correspondence to: Philip David Zelazo, Department of Psychology, Sidney Smith Hall, 100 St George Street, University of Toronto, Toronto, Ontario, Canada. E-mail: zelazo@psych.utoronto.ca

Research on individuals with autism provides a good example. To test the specificity of autistic individuals' difficulty with theory of mind (ToM), Baron-Cohen *et al.* (1986) asked a group of high-functioning children and adolescents with autism to arrange sets of pictures to conform to intentional (i.e. ToM), behavioural, and mechanical stories, and found that the individuals with autism were selectively impaired on the intentional stories, whereas this dissociation was not observed in typically developing 4-year-olds and low-functioning individuals with Down syndrome.

On the basis of this functional dissociation, Baron-Cohen *et al.* (1986) argued that a specific impairment in ToM is responsible for the severe social, communicative, and imaginative challenges that characterize autism and other pervasive developmental disorders (for a recent description of these disorders, see *DSM-IV-TR*; American Psychiatric Association, 2000). They subsequently suggested that in the typical case, ToM is subserved by an innately specified, domain-specific cognitive mechanism—a module (e.g. Baron-Cohen, 1994; Leslie and Thaiss, 1992). For example, Baron-Cohen (1994) argued that ToM exhibits six features (taken from Fodor, 1983) stipulated to be necessary and sufficient for modularity: obligatory firing, rapid speed, dedicated neural architecture, characteristic ontogenesis, characteristic breakdown, and domain specificity.

The modularity of perceptual systems is well established (e.g. Marr, 1982; Fodor, 1983; Treisman and Kanwisher, 1998), but the notion of cognitive modularity (e.g. Leslie and Thaiss, 1992; Baron-Cohen, 1994, 1995; Sperber, 1994; Atran, 1998) challenges traditional conceptions of cognitive architecture and remains controversial (e.g. Karmiloff-Smith, 1992; Quartz and Sejnowski, 1994, 1997; Elman *et al.*, 1996). Nonetheless, bolstered by arguments from evolutionary psychology (e.g. Barkow *et al.*, 1992) and reports regarding remarkably specific consequences of neurological damage in adults (e.g. Anderson *et al.*, 1990; Damasio, 1990; Caramazza and Shelton, 1998), this approach gained considerable attention not only in the study of autism, but also in psychology more generally. For example, Leslie (1994) attempted to explain infants' understanding of physical objects (or mechanical agency) in terms of a theory of body mechanism (ToBy); Blair (1995) suggested that psychopaths fail to inhibit violence because they lack violence inhibition mechanisms (VIMs); and, in addition to a ToM mechanism (ToMM), Baron-Cohen (1994) postulated the existence of an eye direction detector (EDD), an intentionality detector (ID), and a shared attention mechanism (SAM).

Despite this widespread interest in cognitive modularity, however, the description of autistic behaviours in terms of a deficient ToM module can be questioned. First, as a general methodological point, the argument for domain specificity of deficits in ToM is weakened by the difficulty of ruling out the possibility that some (perhaps even slightly) more general deficit can account for poor performance on ToM tasks. Demonstrating domain specificity requires careful cross-domain comparisons (the method of 'fine cuts'; Frith and Happé, 1994), which in turn requires a metric for determining the intrinsic intellectual demands of tasks in different domains. In the absence of such a metric, researchers cannot control for domain-general reasoning requirements; they run the risk of comparing a simple task from one domain to a more complex and difficult task from another and, not surprisingly, finding that group differences emerge only on the complex task.

This problem may apply to previous reports of specificity in autism (e.g. Baron-Cohen *et al.*, 1986; Scott and Baron-Cohen, 1996; Scott *et al.*, 1999) because performance on ToM tasks was compared to performance on non-ToM tasks that did not make equivalent intellectual demands. For example, as Baron-Cohen *et al.* acknowledge, the intentional stories that they used were intrinsically more complex than the behavioural and mechanical ones. Thus, an intentional story such as, 'Boy puts chocolate in box; goes out to play; mum eats chocolate; boy sees chocolate gone,' involves several different events whereas mechanical stories such as, 'Rock on hilltop; rock topples; rolls down hill; knocks tree over,' or 'Rock on hilltop; rock topples; rolls down hill; knocks man over,' involve one (or two) events extended over several scenes. Moreover, both the mechanical and the behavioural stories (e.g. 'Boy puts on trousers; then T-shirt; then shoes; is fully dressed') correspond to canonical scripts (Schank and Ableson, 1978) whereas all of the intentional stories involve exceptions to such scripts.

In any case, however, Baron-Cohen *et al.*'s (1986) original findings regarding the specificity of autistic individuals' deficit in mental-state reasoning were not replicated in several later studies (Oswald and Ollendick, 1989; Prior *et al.*, 1990; Ozonoff *et al.*, 1991). For example, Oswald and Ollendick (1989) found no differences in the Baron-Cohen *et al.* (1986) picture sequencing task between a relatively low-functioning group of individuals with autism and an age- and non-verbal-IQ-matched group of adolescents with mental retardation. Similarly, Ozonoff *et al.* (1991) found no group differences on the intentional stories (i.e. the ToM items) between a relatively large group of high-functioning individuals with autism and a well-matched clinical comparison group that actually performed *better* on each of the mechanical subtests (i.e. the physical causality items) and on one of the two behavioural subtests.

Other reports of the specificity of deficits in autism raise similar questions. For example, reports that individuals with autism show deficits in understanding false beliefs but not false photographs (Leekam and Perner, 1991; Leslie and Thaiss, 1992) may need to be re-interpreted in light of Slaughter's (1998) more recent finding that the false photograph task is also easier than the false belief task for typically developing children. Scott *et al.*'s (1999) finding that individuals with autism perform relatively well on tests of counterfactual-syllogistic reasoning is also open to question because it is unclear whether some of the 'syllogisms' actually required participants to integrate information provided in the premises. For example, participants were told, 'All bananas are pink; John is eating a banana,' and then asked, 'Is the banana pink?' To answer the test question, participants simply had to consult the first premise; the second premise could be ignored. Thus, the test may not assess the ability to make logical inferences. Even more problematic is that the correct response on every trial was 'Yes,' so successful performance could be attributed to a 'Yes' bias, rather than to intact counterfactual reasoning. Leevers and Harris (2000) recently demonstrated that individuals with autism performed poorly on some of the same simple counterfactual syllogisms when 'No' responses were required.

Overall, the evidence for a specific deficit in ToM in individuals with autism is equivocal. Moreover, deficits associated with autism appear to extend beyond ToM into other areas of cognition (e.g. Hermelin and O'Connor, 1970; Prior, 1979; Frith, 1989; Prior and Hoffman, 1990; Ozonoff *et al.*, 1991; Tager-Flusberg, 1991; Hughes and Russell, 1993; McEvoy *et al.*, 1993; Hughes *et al.*, 1994; Frith and Happé, 1994). In particular, individuals with autism consistently perform poorly on tasks designed to assess executive function—the processes involved in the control of thought and action (e.g. Hughes and Russell, 1993; Ozonoff *et al.*, 1991).

Based on these findings, some researchers proposed that autism involves deficits in executive function *in addition* to a deficit in a ToM module (Happé and Frith, 1995; Roth and Leslie, 1998), whereas others suggested that impairments in

ToM derive from more general cognitive or executive impairments (e.g. Ozonoff *et al.*, 1991; Harris, 1993; Hughes and Russell, 1993; Frye *et al.*, 1995; Zelazo *et al.*, 1996a). According to these more domain-general views, ToM tasks are complex social problem-solving tasks that involve several problem-solving phases (e.g. flexible problem representation, planning, rule use) and require general psychological processes (e.g. working memory) in addition to relevant knowledge about the social or psychological domain (Zelazo *et al.*, 1997). Limitations in any aspect of problem solving may lead to impairments in ToM. The finding that performance on executive-function tasks is related to performance on ToM tasks (see Ozonoff, 1995, for a review) supports this suggestion, but further work is required in order to determine more precisely which aspects of executive function are implicated (cf. Griffith *et al.*, 1999).

According to one of these more domain-general theories, the cognitive complexity and control (CCC) theory (Frye *et al.*, 1995; Zelazo and Frye, 1997, 1998; Frye *et al.*, 1998), ToM is related to executive function in both typically and atypically developing individuals because both ToM and measures of executive function involve rule use. In particular, on this account, ToM tasks such as the false belief task (Wimmer and Perner, 1983) require the ability to use higher order rules to consider a perspective from which to reason before making a judgement from that perspective. The use of a higher order, if–if–then rule allows one to integrate two incompatible perspectives into a single system of inferences, so that a person can say (for example), 'I know that the chocolate is now in the green cupboard, but if I'm asked about Maxi, then if the question is where will he look for the chocolate, then the answer is in the blue cupboard.' According to the CCC theory, typical and atypical development both involve domain-general changes in the complexity of the rule systems that people are able to formulate and use. Complexity, which is measured in terms of the number of degrees of embedding in a system of inferences, provides a metric for making cross-domain comparisons. The theory postulates that children first acquire the ability to use a pair of arbitrary rules (typically around 3 years of age), and then acquire the ability to integrate two incompatible pairs of rules into a single rule system via a higher order rule (typically around 5 years).

The CCC theory makes specific predictions about the range of tasks on which someone who fails certain tests of ToM will have difficulty, and these predictions are supported by evidence from typically developing children and individuals with Down syndrome. In an initial study with typically developing preschoolers, Frye *et al.* (1995) created two tasks that assess the use of higher order rules but do not involve the understanding of mental states. One of these tasks, a ramp task, involves understanding of mechanical causality. In this task, children are shown a ramp with two input holes at the top and two output holes at the bottom. A marble inserted at the top will either roll straight down the ramp or across it, depending on the ramp's configuration (which is signaled by a light). In the across configuration, children arguably need to reason as follows: 'If the light is on, then if the marble is inserted on the left, then it will emerge on the right.'

The other task, the dimensional change card sort (DCCS), is more recognizably an executive function task. This task requires the use of rules to sort cards. In the colour–shape version of the DCCS, participants are shown two target cards (e.g. a red car and a blue flower), told two rules for separating test cards by one dimension (e.g. 'All the red ones go here, but all the blue ones go there.'), and then shown test cards (e.g. red flowers and blue cars) all of which would be sorted differently by colour or by shape. After sorting a number of cards by the first dimension, they are told to switch and sort the cards by the other dimension.

Frye *et al.* (1995) found that on these rule-use tasks, 3-year-olds were able to reason from one perspective (i.e. they used a single pair of rules), but could not switch to another incompatible one (i.e. they apparently failed to use a higher order rule for integrating the two incompatible rule pairs); 4- and 5-year-olds performed well. Analogous age-related changes were found for the standard ToM tasks, and there were positive correlations among tasks with age partialled out. This basic finding was replicated (Carlson and Moses, 2001; see also Perner and Lang, 1999), and similar findings were also found among individuals with Down syndrome, who were impaired on both ToM and rule use relative to a mental-age matched comparison group of typically developing children (Zelazo *et al.*, 1996a; see also Charman *et al.*, 1998; Yirmiya *et al.*, 1998). These findings are inconsistent with both (1) the claim that ToM is a domain-specific, modular function (the specificity claim), and (2) the claim that deficits in ToM are unique to individuals with autism (the uniqueness claim; Zelazo *et al.*, 1996a).

OVERVIEW OF THE CURRENT STUDY

The purpose of the current study was to assess the specificity of individual differences in ToM in individuals with autism and autism-spectrum disorders. Twelve severely impaired (Verbal IQ [VIQ]$\leqslant$40) and 10 mildly impaired (VIQ>40) individuals with autism and autism-spectrum disorders were tested on two ToM tasks, two equally complex rule-use tasks (the ramp task and the DCCS), and 4 discriminant tasks, including two measures of short-term memory (the visually cued recall [VCR] task and the central-incidental memory task) and two problem-solving tasks that did not require the use of higher order rules. One of these problem-solving tasks, a multistep multilocation search task (Zelazo *et al.*, 1998), was chosen to be sensitive to individual differences in participants with mental ages in the young preschool range (i.e. between 2 and 3 years). This task, which can be considered a relatively difficult version of the A-not-B task (Piaget, 1954), assesses goal-directed search and requires children to remember the location of a hidden object and use this information to govern their search for the object despite a prepotent tendency to search where the object was found on previous trials. The other problem-solving task, a causal reasoning task that assessed understanding of simple mechanical transformations (Das Gupta and Bryant, 1989), was expected to be more appropriate for participants with higher VMAs (i.e., over 3 years). This task, as administered here, was selected to assess the use of a single pair of rules but not higher order rules (for discussion of this task vis-à-vis CCC theory, see Frye *et al.*, 1996; Zelazo *et al.*, 1996c).

Both mildly and severely impaired participants were tested in order to assess the extent to which any findings generalized across the spectrum of intellectual function. However, we expected that the tasks that relied heavily on verbal instructions (especially, ToM and rule use) might be inappropriate for the severely impaired participants (the majority of whom also had relatively low VMAs; see Method section). For the mildly impaired participants, however, the following predictions were made: If ToM involves modular mechanisms that are independent of the mechanisms that mediate reasoning in other domains, then there is no reason to expect that ToM should be related to rule use on the ramp task and the DCCS.[1] Alternatively, if these tasks all depend on the ability to use higher order rules, then ToM should be related to rule use. Because the purpose of this study was to assess whether the relations among tasks that are obtained with typically developing children are also obtained with an atypical population with established impairments on ToM tasks, and not to provide a precise characterization of autism, we did not include a comparison group. Indeed, our purpose was *not* to identify autism-specific deficits. Rather, we expected that some participants would succeed on the tasks assessing rule use and ToM, and we needed such variation in performance in order to be able to detect possible relations between tasks. Our expectation that some participants would succeed on both types of task was consistent with previous research indicating that a sizable minority of high-functioning individuals with autism with VMAs similar to those in the current study succeed on measures of ToM (e.g. Happé, 1995).

METHOD

Participants

The participants included 37 individuals with some initial diagnosis of autism–spectrum disorders. Fifteen of these individuals were dropped from the study because they (a) refused to be tested on the Peabody picture vocabulary test-revised (PPVT–R; Dunn and Dunn, 1981; $n = 2$), (b) failed to complete the PPVT-R ($n = 1$), or (c) failed to obtain a minimum VMA equivalent score of at least 3 years of age ($n = 12$).[2] Thus, the final group consisted of 22 participants who completed all tasks. Eighteen of these participants had received diagnoses of autism based on *DSM-III-R* (American Psychiatric Association, 1987) criteria by local educational and/or medical teams, and 1 had received a diagnosis of pervasive developmental disorder (PDD). The remaining three cases had not received clear diagnoses but were suspected of having autism-spectrum disorders. (One of these participants had also been diagnosed with attention deficit disorder with hyperactivity, oppositional disorder, and bipolar disorder). Three of the 22 participants had seizure disorders, 1 had a respiratory disorder, and 1 had a speech impediment.

The PPVT-R was administered to obtain an estimate of VMA. The resulting VMA estimates were then used to obtain VIQ estimates ([Mental age/Chronological age] * 100). Individuals with IQ scores above 70 are not typically classified as mentally retarded; those with IQ scores between 40 and 70 are classified as mildly or moderately retarded; and those with IQ scores below 40 are typically classified as severely or profoundly retarded (Zigler and Hodapp, 1986). Accordingly, we divided our sample into two groups—a mildly impaired group and a severely impaired group—using a cutoff VIQ score of 40, although all participants had a VMA of at least 3 years in both groups. Thus, our final sample consisted of 10 mildly impaired participants (all males) and 12 severely impaired participants (8 males and 4 females; see Table 1 for characteristics of mildly and severely impaired participants).

Design

Participants were tested during three separate sessions lasting between 15 and 45 min. All participants received their three sessions within 35 days ($M = 11$ days, S.D. = 10.26 days). In Session 1, participants received the PPVT-R and a comprehension test assessing knowledge of the various colour, shape, and size terms that were to be used in the DCCS. Sessions 2 and 3 comprised two blocks of tasks that were administered in a counterbalanced order (i.e. half of the participants received the first block during Session 2). The first block included (1)

Table 1. Characteristics of the mildly (MI; $n = 10$) and severely impaired (SI; $n = 12$) participants

	Group	
Characteristic	MI	SI
Chronological age (CA; in years)		
Mean	10.30	17.47
Standard deviation	3.43	6.06
Range	7.58–16.17	9.25–27.08
Verbal mental age (VMA; in years)		
Mean	6.15	4.07
Standard deviation	2.79	0.67
Range	3.67–11.08	3.00–5.42
*Verbal IQ (VMA/CA*100)*		
Mean	59.60	25.59
Standard deviation	18.51	8.14
Range	41.51–98.86	14.20–36.97

an explicit false-belief task, (2) an unexpected-contents ToM task (including appearance-reality questions, a representational-change question, and a false-belief question), (3) the DCCS, and (4) the ramp task. The order of tasks within this block was counterbalanced across participants such that participants did not receive the two ToM tasks in succession, nor did they receive each of the two rule-use tasks in succession. Thus, eight orders of task presentation were possible for this block of tasks, and participants were randomly assigned to one of these orders.

The second block of tasks included: (1) the multistep multilocation search task, (2) the visually cued recall (VCR) task, (3) a central-incidental memory task, and (4) the causal-inference task, in addition to other tasks that will not be discussed in this paper. The order in which these tasks were administered was determined randomly for each participant.

Procedure

Peabody picture vocabulary test-revised (PPVT-R; Dunn and Dunn, 1981)

The PPVT-R was administered in the standard fashion (see Dunn and Dunn).

Comprehension test of colour, shape, and size terms

The inclusion of this test ensured that performance on the DCCS was not compromised by difficulty identifying the correct cues for each dimension. Participants' receptive knowledge of colour, shape, and size terms was assessed by asking them to select a card from three arrays of items representing colour, shape, and size terms, respectively (e.g. 'Can you give me the red one?'). Terms that were tested included *red, blue, green*, and *yellow* for colour; *car, flower, boat, rabbit, circle*, and *triangle* for shape; and *big* and *small* for size. All of the participants except for one selected all colours and shapes correctly, and red and blue cars and flowers were used in the DCCS for these participants. The remaining participant failed to select the correct items for colour, but selected the correct items for size. Thus, big and small cars and flowers were used in the DCCS for this participant.

Theory of Mind Tasks

Explicit false-belief task

In the explicit false-belief task (Wellman and Bartsch, 1988; change-of-location paradigm), participants were told that a kitten was in one of two locations. They were then told explicitly that a character believed that the kitten was in the other location. While illustrating the events of the story with pictures, the experimenter told participants,

> Jane and her kitten are playing in the living room. But Jane decides that she's tired. So she goes into her bedroom to take a nap and leaves her kitten in the living room. While Jane is sleeping, the kitten decides it's hungry. So it goes into the kitchen to have a snack. Jane wakes up and wants to go find her kitten. Jane's kitten is *really* in the kitchen. Jane *thinks* her kitten is in the living room. Where will Jane look for her kitten? Where is the kitten really?

Story pictures remained visible throughout the task and while participants were asked a belief question (i.e. 'Where will Jane look for her kitten?') and a reality question (i.e. 'Where is the kitten really?'), the order of which was counterbalanced across participants. They could respond either verbally or manually, by pointing to one of the story pictures.[3] Performance was scored as in Zelazo *et al.* (1996a). Participants received a score of 0 if they answered the reality question incorrectly (regardless of how they did on the belief question), they received a score of 1 if they answered only the reality question correctly, and they received a score of 2 if they answered both questions correctly.[4]

Unexpected-contents theory of mind task

The questions used in this task were adapted from previous studies (e.g. Flavell *et al.*, 1983; Perner *et al.*, 1987; Gopnik and Astington, 1988). A crayon box containing sticks was used as the 'deceptive' item. Pictures of sticks and crayons were first placed in front of participants so that they could respond either verbally or by pointing to a picture. Participants were then shown the crayon box and asked to identify it: 'What's in the box?' The box was then opened to reveal that it contained sticks. Participants were asked again, 'What's in the box?' The box was then closed and placed out of reach, and participants were asked the ToM questions (including false belief, representational change, and appearance-reality) in a counterbalanced order. For the false-belief question, participants were shown a puppet and asked, 'I'm going to ask you a couple of questions and Pat wants to help. Pat is seeing this box for the first time. She's never opened up the box and looked inside, so she's seeing it for the very first time. What does Pat think is in the box?' For the representational-change questions, participants were asked, 'When I brought out this box at first, *before* we opened it, what did you think was in the box?' The order in which the appearance and reality questions were presented was counterbalanced across participants: 'What does it look like is in the box?' [appearance question] and 'What's really in the box?' [reality question].

As with the explicit false-belief task, participants received a score of 0 if they answered the reality question incorrectly, irrespective of how they answered the other questions. If they answered the reality question correctly, they received a score of 1 and one additional point for each of the other questions that they answered correctly (i.e. false belief, representational change, and appearance). Thus, scores on this task ranged from 0 to 4.

Rule-Use Tasks

Dimensional change card sort (DCCS)

On the DCCS (Frye *et al.*, 1995; Zelazo *et al.*, 1996b), participants were shown two target cards (i.e. a red car and a blue flower) that were affixed to one of two sorting trays. They were then told two pre-switch rules (e.g. colour rules) for sorting a series of test cards that all matched one target card on one dimension and the other target card on the other dimension (i.e. test cards were blue cars and red flowers): 'We're going to play the colour game. In the colour game, put all the red ones here and all the blue ones there.' The experimenter modelled two test cards (one of each type) and then gave participants five trials with the pre-switch pair of rules. On each trial, the experimenter repeated the rules for sorting test cards and labeled the test card by both dimensions (e.g. 'Here's a red flower. Where does it go?'). Participants received test cards in a pseudorandom order such that no more than two test cards of the same type were presented consecutively. Participants were not told whether or not they sorted correctly. After this pre-switch phase, the experimenter explained the post-switch rules (e.g. shape rules), which were presented in the same manner as the pre-switch rules except that participants were required to sort test cards according to the alternate dimension: 'Now, we're not going to play the colour game any more, we're going to play the shape game. In the shape game, put all the cars in this box and put all the flowers in that box.' Participants received five post-switch trials with test cards presented in the same type of pseudorandom order used during the pre-switch phase.

Approximately half of the participants received the colour rules first, and the other half received the shape rules first. Performance was scored as in previous studies (e.g. Zelazo *et al.*, 1996a). Participants were considered to have passed a phase if they sorted correctly on at least four of the five trials or if they sorted correctly on the last three trials within that phase. For each phase that they passed, they were given a score of 1. Thus, if they passed both phases (indicating use of 2 incompatible pairs of rules) they received a score of 2, if they passed only one phase (indicating use of a single pair of rules) they received a score of 1, and if they failed both phases (indicating failure to use even a single pair of rules) they received a score of 0.

Ramp task

In the ramp task (Frye *et al.*, 1995, 1996), participants were presented with a wooden ramp with two input holes at the top and two output holes at the bottom. Depending on the configuration of the ramp (signaled by a light), marbles either rolled straight down the ramp or followed diagonal paths. The operation of the ramp was demonstrated, and participants were shown how a marble rolled in each configuration (straight and across).

Once participants were shown how the ramp operated in both configurations (with the post-switch rules always demonstrated first), the ramp was covered with an opaque cover. In a pre-switch phase, the ramp was presented in one configuration (light on or off) and participants were given five trials in which they were asked to predict where a marble would exit at the bottom of the ramp given a particular input hole. For example, they were told on each trial, 'When the light is on the marbles roll across. So if I put a marble here [pointing to one of the input holes], where will the marble come out?' Input holes were varied in the same pseudorandom manner as in the DCCS. Participants were then given five post-switch trials in which they had to make opposite predictions (i.e. the ramp was presented in the other configuration). Rules were repeated on each trial in both the pre-switch and post-switch phases and participants were explicitly told to switch between phases. Throughout the task, participants were shown two pictures: one depicting two black parallel lines and the other depicting an 'X'-shaped figure. The experimenter always pointed to the relevant picture when stating the rules (i.e. the experimenter said, 'When the light is on, the marbles roll across', and pointed to the 'X'). This was done to ensure that participants did not fail because they did not understand the words 'across' or 'straight'.

Approximately half of the participants received the straight rules first, whereas the other half received the across rules first. Performance on this task was scored in the same manner as in the DCCS. That is, participants received a score of 1 for each phase in which they made correct predictions on at least four of the five trials or on the last three trials. Unlike the DCCS, in which participants who use only a single pair of rules typically fail the post-switch phase, in the ramp task, participants often perseverate on the straight rules regardless of the order in which they are presented (i.e. they have a pre-experimental bias to use the straight rules). Thus, in the ramp task, it is typically the across configuration that requires the use of a higher order rule.

Discriminant Tasks

Multistep multilocation search task

Participants were familiarized with a hiding apparatus devised by Zelazo *et al.* (1998; see Figure 1). The hiding apparatus consisted of a 55 cm × 60 cm × 30 cm wooden box with a removable top and a side door through which the experimenter could hide rewards (e.g. cereal, chocolate, or a sticker). An opaque front door could be raised to reveal the contents of the apparatus, or it could be lowered (to within 3 cm of the base of the box) to conceal them. A transparent plexiglass front door was also employed. Unlike the opaque door, this door could be lowered completely thus preventing participants from reaching during hiding.

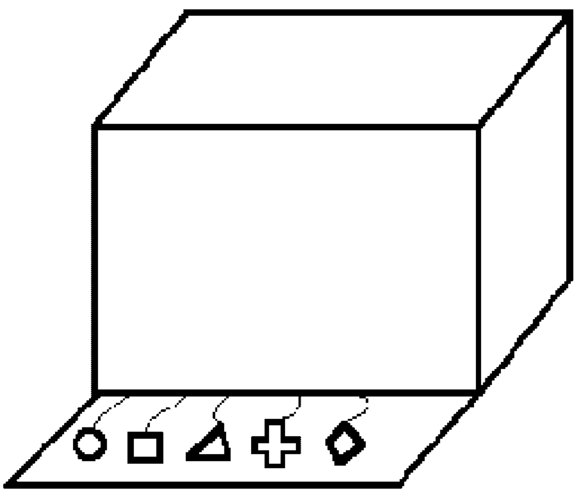

Figure 1. Participants' view of the multistep multilocation hiding apparatus after the foam barrier has been removed and the tray has been pulled out.

On the base of the box was a sliding tray that extended to a maximum of 20 cm from the door. When fully extended, this tray exposed five equidistant velcro swatches to which cardboard symbols (about 2 cm^2) could be affixed. The swatches were separated by 10 cm. Each symbol was connected by a 45 cm string to a transparent plastic bag. The bags were placed in the back of the box and were used to contain the rewards. A 50 cm×10 cm×7 cm foam barrier was placed in front of the tray during hiding.

In a warm-up phase, only one hiding location (the centre) was used, and participants were taught the steps for retrieving the reward: They were shown that to retrieve the reward, they needed to remove the foam barrier, pull out the tray, and pull on the string to obtain a reward. Once participants could correctly retrieve the reward in the warm-up phase, they received the task proper, which involved the same retrieval processes as in the warm-up phase except that five potential hiding locations instead of one were used. The reward was conspicuously hidden at one of the locations (except for the centre location; Location 1), which was randomly determined for each participant. Participants were asked to search for the reward. After they successfully retrieved the reward on five consecutive trials at Location 1, the reward was conspicuously hidden at one of the three remaining locations (Location 2), which was also randomly determined. Participants searched at Location 2 until they successfully retrieved the reward at that location. Dependent measures included the number of trials that participants required in order to search correctly on five consecutive trials at Location 1 (the minimum possible score was 5) and the number of trials participants required to search correctly at Location 2 (the minimum score was 1). In both cases, fewer trials indicated better performance.

Visually cued recall (VCR) task

On this new memory-span task, the experimenter introduced participants to a puppet ('Pat') and presented them with 10 different 3×4 arrays of 12 pictures, one at a time. The experimenter showed participants the first array and said:

> This is my friend Pat. Pat likes certain things very much. I'm going to show you pictures of things that Pat likes and when I finish showing you pictures of things that she likes, I want you to point to them for me. But you can only point to things that Pat likes. OK? See, on this poster, Pat likes the tricycle [simultaneously pointing to the item]. Can you point to the one that Pat likes?

On the next array, the experimenter pointed to two items and required that participants point to both (pointing in the same serial order was not required). On each subsequent trial, a new array was presented and participants were required to remember one more item than in the preceding array. The task was discontinued when participants failed two consecutive trials (or when they reached a maximum of 10 items). The arrays were presented in the same order, and the items that were selected by the experimenter were the same for all participants. Scores on this task consisted of the number of items that participants identified on their last correct trial (i.e. Trial 10 or the trial that immediately preceded the two consecutive incorrect trials).

Central-incidental memory task

This task, adapted from Hagen (1967) and Burack and Zigler (1990), had two components, a central-memory component and an incidental-memory component. The central-memory component assessed short-term memory for target items, and performance on this component reflected appropriate selective attention. In contrast, the incidental-memory component assessed memory for incidental items. In the central-memory component, participants were presented with five cards, each of which included a picture of an animal and a picture of a food item (i.e. rabbit–cheese, horse–ice cream cone, bird–carrot, cow–crackers, and dog–apple). The experimenter told participants to pay attention only to the pictures of the animals and then asked them to name each of the animals (only). If participants did not produce a unique label for each animal, the experimenter told them a label and a short description of the animal.

On each trial, two to five cards were placed face down side-by-side in front of participants and each was briefly turned over one at a time and then placed face down again. Participants were then shown a picture of a target animal and asked to find the upside-down card that included that target animal. Participants first received two practice trials with only two cards (the target and one distracter card). On each of these practice trials, the experimenter provided participants with feedback on their performance, and the trials were repeated if participants failed to find the target item on that trial. The experimenter then administered three test trials, each with two, three, and four distracters present (i.e. a total of nine trials). Participants' score on this component of the task was based on the total number of test trials on which they responded correctly.

When the central-memory component of the task was completed, the experimenter presented the incidental-memory component. In this component, participants were shown cards that contained only pictures of the animals that appeared in the central-memory component. They were then presented 10 pictures of food items (5 that appeared in the central-memory component and 5 new items) and asked to match each food item with the animal with which it was paired in the central-memory component. Scores on the incidental-memory component consisted of the number of correct matches.

Causal-inference task

In this task, adapted from Das Gupta and Bryant (1989), participants were shown an array of five pictures. Two pictures were isolated from the other three pictures. These two pictures depicted the same item in two different states. For example, one pair of pictures consisted of a blank piece of paper and a piece of paper with writing on it. Participants were asked to consider the other three pictures (e.g. matches, pencil, and telephone) and select the object that most likely caused the first picture (i.e. the blank piece of paper) to become like the second (i.e. the piece of paper with writing). The experimenter said, 'What did I use in these three pictures down here to make that piece of paper [the experimenter pointed to the first picture in the pair] become like that piece of paper [the experimenter pointed to the second picture in the pair]? What did I use in these three pictures?'

Participants were administered one demonstration trial and five test trials (the test trials were presented in a random order to each participant). On four of the trials (including the demonstration trial), participants were required to find an object that changed the item from being in its canonical state to being in a non-canonical state (e.g. the blank piece of paper to the paper with writing). On the two remaining trials, the transformation occurred from a non-canonical to a canonical state (e.g. lighted candle to unlighted candle, for which participants had to select among a blow dryer, scissors, and keys). Scores on this task consisted of the number of trials on which participants correctly identified the causal object.

Scoring

Performance was scored by a secondary experimenter who was present during testing. In order to ensure scoring reliability, the primary experimenter rescored all tasks from videotape. However, due to the inaccessibility of videotape equipment at certain testing locations, 8 participants were not videotaped.

RESULTS

To obtain reliable estimates of performance on the rule-use and ToM tasks, composite scores were derived for both types of task. As mentioned previously, on both the DCCS and the ramp task, participants were given a score of 1 for any phase that they passed; otherwise they were given a score of 0. Each of these tasks included two phases (a pre- and post-switch phase), so participants could receive scores of 0, 1, or 2 on each task. Thus, the rule-use composite scores ranged from 0 to 4. ToM composite scores were based on the explicit false-belief task and the unexpected-contents ToM task, and ranged from 0 to 6. The overall correlation between the two rule-use tasks was, $r = 0.55$, $p < 0.01$, and the correlation between the two ToM tasks was $r = 0.43$, $p < 0.05$.

Preliminary Analyses

Descriptive statistics for all measures were examined in order to assess whether the measures were suitable for both mildly and severely impaired participants. Separate t tests and F_{max} tests were computed for all measures to determine whether there were any group (mildly impaired vs severely impaired) differences and to determine whether there were any ceiling or floor effects for either group on any measure. As can be seen in Table 2, the means or standard deviations of each group differed significantly from each other on most measures (except for the central-memory component of the central-incidental memory task). Moreover, the mildly impaired participants had larger mean scores and larger standard deviations than the severely impaired group on most measures (i.e. the rule-use composite, the ToM composite, the VCR, the central- and incidental-memory components of the central-incidental memory task, and the causal-inference task) because the severely impaired group's scores tended to be restricted toward the lower end of the possible scores (see the ranges in Table 2 for each group as a function of measure). The only measures on which the severely impaired group had larger mean scores and larger standard deviations than the mildly impaired group were search at Location 1 and search at Location 2 on the multistep multilocation search task (where lower scores were indicative of better performance). On these two measures, the mildly impaired group generally required fewer trials to find the hidden object at both locations. Thus, in this instance, the mildly impaired group performed near ceiling, whereas the severely impaired group did not perform near floor.

Preliminary analyses also revealed a main effect of phase on the DCCS, $F(1, 19) = 11.4$, $p < 0.005$, indicating that participants performed better on the pre-switch phase than on the post-switch phase. (Only 21 participants were included in this analysis because one participant received size rules rather than colour

Table 2. Descriptive statistics for mildly (MI) and severely impaired (SI) participants, including means, standard deviations, ranges, and group mean and variance differences

Measure	Mean	Standard deviation	Range	*T*-test value	*T*-test prob.	F_{max} prob.	
Theory of mind composite (min=0, max=6)							
MI	2.50	1.72	0–5	1.80	0.09	1.88	0.32
SI	1.36	1.25	0–3				
Rule-use composite (min=0, max=4)							
MI	3.00	0.94	2–4	2.75	0.01**	1.10	0.87
SI	1.92	0.90	1–4				
Multistep multilocation search task (Loc. 1) (min=5, max=20)							
MI	5.10	0.32	5–6	−1.25	0.23	31.14	0.001**
SI	5.75	1.76	5–11				
Multistep multilocation search task (Loc. 2) (min=1, max=10)							
MI	1.10	0.32	1–2	−1.25	0.23	31.14	0.001**
SI	1.75	1.76	1–7				
Visually Cued Recall Task (min=0, max=10)							
MI	4.30	2.41	1–9	2.25	0.04*	1.39	0.59
SI	2.17	2.04	0–6				
Central-memory component (min=0, max=9)							
MI	5.90	2.51	2–9	1.51	0.15	1.88	0.32
SI	4.50	1.83	2–8				
Incidental-memory component (min=0, max=5)							
MI	1.60	2.07	0–5	1.35	0.20	7.04	0.003**
SI	0.67	0.78	0–2				
Causal-inference task (min=0, max=5)							
MI	3.40	1.51	1–5	2.35	0.03*	2.58	0.14
SI	2.17	0.94	1–4				

Note. The minimum (min) and maximum (max) scores that could be obtained on a given measure appear in parentheses after the name of each measure. The degrees of freedom for the t tests were 20. For the F_{max} test, they were 9 and 11. However, for measures with significant F_{max} values, t-test values and degrees of freedom corrected for unequal variances were used for testing differences between means.
*$p < 0.05$;**$p < 0.01$.

rules; see Procedure). There was no effect of the order in which rule pairs (colour or shape) were presented on this task, and order did not interact with phase. In contrast, on the ramp task, there was no effect of phase and no effect of order, but there was an interaction between phase and order, $F(1, 20) = 12.5$, $p < 0.005$, indicating that participants performed better with the straight rules than the across rules regardless of the order in which these rules were presented.

Correlations Among Tasks

We predicted that performance on non-mentalistic rule-use tasks (i.e. the DCCS and the ramp task) would be related to performance on ToM tasks (i.e. the explicit false-belief task and the unexpected-contents ToM task), whereas

Table 3. Correlations for mildly impaired participants

	VMA	CA	VIQ	ToM	Rule	VCR	CENT	INCI
VMA	—							
CA	0.80**	—						
VIQ	0.59	0.01	—					
ToM	0.50	0.38	0.24	—	—			
Rule	0.58	0.32	0.48	0.82**	—			
VCR	0.34	−0.12	0.75*	0.23	0.59	—		
CENT	0.08	−0.30	0.49	−0.06	0.37	0.69*	—	
INCI	0.24	0.28	−0.02	0.19	0.29	0.05	0.31	—
CAUSAL	0.84*	0.70**	0.56	0.34	0.47	0.48	0.07	0.34

Note. Correlations were not calculated for the multistep multilocation search task for the mildly impaired participants because only 1 participant required more than 5 trials at Location 1, and only 1 (other) participant required more than 1 trial to search correctly at Location 2. VMA=verbal mental age; CA=chronological age; VIQ=verbal IQ; ToM=composite score for ToM tasks; Rule=composite score for rule-use tasks; VCR=score on the Visually Cued Recall task; CENT=score on central-memory component of the central-incidental memory task; INCI=score on incidental-memory component of the central-incidental memory task; CAUSAL=score on the causal-inference task.
$^{*}p<0.05$; $^{**}p<0.01$.

Table 4. Correlations for severely impaired participants

	VMA	CA	VIQ	ToM	Rule	Loc.1	Loc.2	VCR	CENT	INCI
VMA	—									
CA	0.35	—								
VIQ	0.04	−0.89**	—							
ToM	−0.05	0.21	−0.29	—						
Rule	0.02	−0.58	0.64*	−0.40	—					
Loc. 1	−0.44	−0.08	−0.20	0.44	−0.41	—				
Loc. 2	−0.47	−0.14	−0.13	0.53	−0.41	0.97**	—			
VCR	0.07	−0.14	0.01	−0.19	0.16	0.01	−0.11	—		
CENT	−0.02	0.52	−0.61*	0.18	−0.30	−0.07	−0.07	0.22	—	
INCI	0.24	0.41	−0.38	0.17	−0.56	0.07	0.07	−0.25	0.32	—
CAUSAL	0.03	−0.03	0.07	−0.44	0.45	−0.41	−0.41	−0.02	0.32	0.08

Note. VMA = verbal mental age; CA = chronological age; VIQ = verbal IQ; ToM = composite score for ToM tasks; Rule = composite score for rule-use tasks; Loc. 1 = number of trials to criterion for Location 1 on the multistep multilocation search task; Loc. 2 = number of trials to criterion for Location 2 on the multistep multilocation search task; VCR = score on the Visually Cued Recall task; CENT = score on central-memory component of the central-incidental memory task; INCI = score on incidental-memory component of the central-incidental memory task; CAUSAL = score on the causal-inference task.
$^{*}p<0.05$; $^{**}p<0.01$.

performance on discriminant tasks (i.e. the multistep multilocation search task, the VCR, the central-incidental memory task, and the causal-inference task) would not be related to performance on ToM. The overall correlation across all participants between the ToM composite and the rule-use composite was not significant, $r=0.40$, $p<0.07$, although it approached significance. However, because of the apparent floor performance of the severely impaired group on the rule-use and ToM tasks, it is more appropriate to calculate correlations separately for the mildly and severely impaired groups.

Table 5. Verbal mental age (VMA) and chronological age (CA) and composite scores on the theory of mind (ToM) and rule-use (Rule) tasks for each participant by group

Participants	VMA	CA	Rule	ToM
Mildly impaired participants				
1	3.67	8.42	2	0
2	6.50	13.67	2	1
3	3.75	8.25	2	1
4	4.00	7.58	2	3
5	7.83	7.92	3	1
6	4.08	9.83	3	2
7	4.17	8.00	4	4
8	6.00	7.67	4	4
9	11.08	15.50	4	4
10	10.42	16.17	4	5
Severely impaired participants				
1	3.42	24.08	1	0
2	4.17	27.08	1	2
3	5.42	25.50	1	2.33[a]
4	3.00	12.50	1	3
5	4.67	14.83	2	0
6	3.42	9.25	2	0
7	4.50	18.75	2	1
8	4.58	18.92	2	1
9	4.17	11.67	2	3
10	3.58	21.67	2	3
11	3.75	14.00	3	1
12	4.17	11.42	4	0

Note. The possible values for the rule-use composite ranged from 0 to 4, whereas those for the ToM composite ranged from 0 to 6.
[a] This participant refused to answer the false-belief question in the unexpected-contents ToM task. Thus, an average score out of 4 was derived on the basis of the participant's performance out of 3 on the unexpected-contents ToM task.

Accordingly, the results of the correlational analyses are presented separately in Tables 3 and 4 for the mildly and severely impaired groups, respectively. For the mildly impaired participants, performance on the rule-use tasks was significantly related to performance on ToM tasks, $r = 0.82$, $p < 0.01$,[5] and as predicted, no other measures correlated with performance on ToM tasks.[6] Additionally, performance on the VCR correlated with both VIQ ($r = 0.75$, $p < 0.05$) and the central-memory component of the central-incidental memory task ($r = 0.69$, $p < 0.05$), and performance on the causal inference task correlated with both VMA ($r = 0.84$, $p < 0.05$) and chronological age ($r = 0.70$, $p < 0.01$).

For the severely impaired participants, rule use did not correlate with the ToM composite ($r = -0.40$, $p > 0.10$),[7] likely as a result of the restricted range of scores for this group (see Table 5, which presents the rule-use and ToM composites for individual participants in each group). Instead, for this group, performance on rule use tasks correlated with VIQ ($r = 0.64$, $p < 0.05$). Moreover, VIQ was negatively related to both chronological age ($r = -0.89$, $p < 0.01$) and the central component of the central-incidental memory task ($r = -0.61$, $p < 0.05$). Finally, search at Location 1 on the multistep multilocation search task was highly related to search at Location 2 ($r = 0.97$, $p < 0.01$).

DISCUSSION

In the present study, we assessed the domain specificity of individual differences in ToM in people with autism-spectrum disorders by comparing performance on ToM tasks to performance on two rule-use tasks that do not involve reasoning about mental states, but are correlated with ToM in typically developing preschoolers (Frye *et al.*, 1995). One of these tasks, the ramp task, required participants to reason about mechanical causation, whereas the other task, the DCCS, required them to appreciate and act upon a stipulated behavioural regularity. The CCC theory provided a complexity metric for selecting these two tasks, and hence, for making these cross-domain comparisons. A series of discriminant tasks was also administered. These tasks assessed various aspects of cognitive function (short-term memory, goal-directed search, causal reasoning), but unlike the ToM tasks and the rule-use tasks, they were not designed to involve the use of higher order rules to integrate two pairs of incompatible rules.

Preliminary analyses revealed that most of the tasks were too difficult for the severely impaired participants; floor effects were found for ToM, rule use, and several of the discriminant tasks. Not surprisingly, given these floor effects, ToM was unrelated to rule use in this group. The task that best captured meaningful individual variation in the severely impaired participants was the multistep multilocation search task: for the severely impaired participants, performance on the pre-switch phase of this task was highly related to performance on the post-switch phase. The multistep multilocation search task was too easy for the mildly impaired participants (80% of whom performed at ceiling), but most of the other tasks were developmentally appropriate for this group. In this group, the correlation between ToM and rule use was high. Performance on the discriminant tasks was unrelated to ToM or to rule use in this group, although performance on the VCR was related to performance on the central-memory component of the central-incidental memory task.

It remains unclear whether the severely impaired participants performed more poorly on most tasks because of their lower VIQ or because of their lower VMA. A greater percentage of severely impaired than mildly impaired participants had VMAs between 3 and 4 years of age (42% vs 20%), and typically developing 3-year-olds perform more poorly on rule use and ToM tasks than 4-year-olds. However, low VIQ might lead to cognitive impairments in and of itself. For example, for a given VMA, individuals with autism with lower VIQs will be older than their counterparts with higher VIQs, introducing the possibility of cohort effects (e.g. poorer services, greater probability of being institutionalized, etc.). In addition, measures of MA may be less reliable indices of level of function among individuals with low IQs (Zelazo *et al.*, 1996a; Charman *et al.*, 1998). In any case, future research needs to be conducted with lower functioning individuals to determine whether low IQ leads to deficits above and beyond those attributable to VMA (Zelazo *et al.*, 2001).

Previous demonstrations of the specificity of ToM in individuals with autism were based on differential patterns of success and failure on tasks that differed not only in content domain (e.g. ToM vs mechanical causality), but also in complexity. By treating participant group as an independent variable (though it is really a correlate), researchers demonstrated what is effectively a single functional dissociation: Autism affects one task, ToM, but not another (e.g. sequencing of pictures according to a mechanical story). However, the limitations of this approach are well known: single functional dissociations cannot support inferences regarding the independence of functional processes (Dunn and

Kirsner, 1988). By using a complexity metric to select two non-ToM rule use tasks (i.e. the DCCS and the ramp task) hypothesized to make equivalent intellectual demands, we were able to take a different approach and provide a relatively sensitive test of the specificity of ToM. Because these tasks and the ToM tasks were chosen to be equally difficult—both formally and empirically in typically developing children—they arguably differed only in content. Moreover, instead of attempting to determine whether individuals with autism showed a pattern of specific impairments (e.g. impairments in ToM but not rule use), we took an individual-differences approach (which would not be informed by the presence of a comparison group). As individuals with autism are impaired on ToM (e.g. Baron-Cohen *et al.*, 1985), the question we posed was whether or not this impairment was related to other aspects of cognitive and executive function.

The finding that performance on ToM and rule use tasks are correlated for the mildly impaired individuals challenges the hypothesis of a domain-specific, modular impairment in ToM, and it adds to previous reports that performance on ToM is related to tasks assessing executive function (e.g. Frye *et al.*, 1995; Hughes, 1998; see Perner and Lang, 1999, for a review). Various explanations of the relation between executive function and ToM have now been proposed. Whereas some authors (e.g. Ozonoff *et al.*, 1991; Russell *et al.*, 1991; Hughes, 1998) suggest that poor performance on ToM tasks can be attributed to failures of executive function, or related aspects of cognition such as working memory (e.g. Gordon and Olson, 1998), others (e.g. Perner *et al.*, 1999) suggest that ToM is required for the control of action.

The CCC theory integrates these two approaches, and provides a relatively precise explanation of this integration. On this account, poor performance on ToM tasks can be attributed to failures of executive function because ToM can be considered an example of executive function in a particular content domain (Zelazo and Mueller, in press). Moreover, ToM is required for executive function insofar as ToM involves reflection on one's own mental states (Frye and Zelazo, in press). Performance on particular measures of executive function will be related in both typically and atypically developing individuals when they require rule use of comparable complexity. The particular measures of executive function (including, on this account, ToM) that 3-year-olds find difficult are those that require the use of a higher order rule to integrate two incompatible perspectives into a single system of inferences. This system of inferences allows one first to select a perspective and then to make a judgment from that perspective. For example, in a representational change task (Gopnik and Astington, 1988), one must be able to reason flexibly about one's own past knowledge vis-à-vis one's appraisal of the current situation. To do so, one must say something like: 'There are sticks in the box, not crayons, but I'm being asked about before I saw that, so the answer is crayons, not sticks.' That is, one must formulate rules in natural language (based on both the task demands and one's knowledge about behaviour and mental states) that allow one to access and focus on particular pieces of knowledge at the time of responding. In contrast to typically developing 5-year-olds, 3-year-olds and individuals with autism may perform poorly on these ToM tasks because they are unable to integrate the incompatible perspectives into a higher order rule that allows them to select the appropriate perspective from which to reason. Failure to consider the perspective as well as the type of judgement will result in perseveration on the more salient perspective.

The results of the current study suggest that ToM is indeed related to these measures of rule use in a group of mildly impaired individuals with autism-spectrum disorders. Because the results are correlational, they are open to a variety of interpretations. Performance on ToM may depend on executive function as assessed by the measures of rule use (e.g. Ozonoff *et al.*, 1991; Russell *et al.*, 1991; Hughes, 1998), ToM may be required for performance on the measures of rule use (e.g. Perner *et al.*, 1999), or performance on both ToM and rule use may depend on common aspects of cognitive function, such as working memory (e.g. Gordon and Olson, 1998), counterfactual reasoning (e.g. Riggs and Peterson, 2000), spontaneous generativity (e.g. Peterson and Bowler, 2000), or relational complexity (Halford *et al.*, 1998). Our own favoured interpretation, however, is that important changes in rule use, including the acquisition of the ability to use a higher order rule, underlie performance in a variety of domains even in atypical development (see Zelazo and Jacques, 1996, for a review). This approach characterizes the underlying cognitive processes involved in performance on ToM tasks, but it does not discount the importance of domain-specific knowledge. Rather, the ability to use higher order rules is taken to be necessary, but not sufficient, for reasoning about conflicting epistemic states. In addition to rule use, these tasks require general knowledge about the mind in order to support the ad hoc inferences demanded by the task. In cases of atypical development, opportunities for acquiring domain-specific knowledge may be limited, and the lag between the acquisition of underlying reasoning skills and the ability to apply them in any particular domain (e.g. ToM) may be relatively pronounced (Hodapp and Burack, 1990).

The CCC approach is not without its critics. Perner (2000; see also Perner *et al.*, 1999), for example, suggests that the rules identified in our analysis of ToM performance 'cannot be the rules that children bring to bear on the task, because these rules could only be known after a practice run or as a result of the child having figured out the problem' (p. 382). Our claim, however, is not that people must learn these rules, but that they must formulate these rules in an ad hoc fashion—basically, they need to talk their way through the problem in a way that allows them to access the appropriate piece of knowledge. Developmental constraints on the complexity of one's rule formulations determine task difficulty.

Another objection to CCC theory concerns the apparently arbitrary way in which task complexity is analyzed. Perner (2000) provides an alternative analysis of the false belief task according to which it requires a simple pair of rules, rather than a higher order rule for integrating two incompatible pairs of rules. It should be noted, however, that CCC theory does not attempt to provide a logically necessary analysis of the false belief task. Instead, it generates an empirical hypothesis regarding the reasoning processes that underlie performance on the task. Any given 2-choice discrimination, including the false belief task (i.e. false belief response vs reality response), is amenable to analysis in terms of a simple pair of rules. However, the CCC theory holds that the psychological perspectives identified in this task (i.e. the child's correct perspective vs the other person's false perspective) serve naturally as setting conditions for a higher order rule. Thus, the empirical claim is that when solving the task, one first determines from which perspective to reason, and then determines which judgement to make from that perspective. This claim receives empirical support from the current findings, as well as findings from previous research (e.g. see Frye *et al.*, 1998).

The current findings should be treated as preliminary, however, as they are based on a relatively small sample. Although the findings from the mildly impaired group were not compromised by floor and ceiling effects, the findings from the severely impaired group were. Accordingly, the current study raises important questions concerning the role of cognitive complexity in both typical and atypical development, but future research with a larger sample size, additional measures, and participants with a wider range of intellectual function, will be required to answer these questions.

CONCLUSION

An assumption of domain specificity invites psychologists to fractionate the field and misconstrue cognitive constructs by considering them outside of the broader context of reasoning and action (Jacques and Zelazo, 1994). This tendency, in turn, may lead to a proliferation of narrow explanations that could obscure the psychological similarities across a range of situations. Because CCC theory provides a metric for comparing performance on tasks from different domains, it permits a legitimate test of the specificity claim regarding autism and ToM. The results of this test challenge the notion of a domain-specific impairment in ToM in individuals with autism, and instead indicate that poor performance on ToM in these individuals can be attributed to a more general difficulty—potentially that of using higher order rules in order to integrate two incompatible perspectives into one rule system. Further research is still necessary, but the current findings encourage us to consider both typical and atypical cognitive development as a complex epigenetic process involving domain general changes in addition to the acquisition of specific knowledge.

ACKNOWLEDGEMENTS

Part of this work was presented at the biennial meeting of the International Society for the Study of Behavioural Development (ISSBD), Quebec City, Canada, August, 1996. Preparation of this article was supported by grants from the Natural Sciences and Engineering Research Council of Canada (PDZ), the Social Sciences and Humanities Research Council of Canada (JAB), and the Fonds pour la Formation de Chercheurs et l'Aide a la Recherche du Quebec (JAB).
We thank Drs Michael Weiss, Carol Rahn, and Cosmina Vaccarino for their help in recruiting participants for this study. We are also grateful to M. Aikens, O. Amitay, J. Boseovski, V. Huta, G. Iarocci, N. Kirkham, B. Randolph, M. Scharer, and T. Solomon for assistance with data collection and scoring. Finally, we wish to thank the staff, parents, and participants from the following schools and centres: Autism Consultation Services (St Catherine's, Ontario, Canada), Autism Society of Ontario, Niagara Chapter (Niagara Falls, Ontario, Canada), Giant Steps (Fairfield, Connecticut), Giant Steps (Montreal, Quebec, Canada), the Muki Baum Association (Toronto, Ontario, Canada), and the Summit School (Montreal, Quebec, Canada).

Notes

1. A similar point holds for the 'theory theory,' which predicts that 'individual children's progress in one domain will be quite independent from their progress in another' (Gopnik and Wellman, 1994, p. 287).
2. This cutoff was designed to ensure that participants understood the basic task instructions. Use of this cutoff also served to eliminate participants who were simply unable or unwilling to respond by pointing to the experimenter's

questions, thereby reducing the risk that a participant would perform poorly on the other tasks because of a pre-existing difficulty in pointing.

3. On all tasks, it was possible for participants to answer questions either verbally or non-verbally. Most participants tended to answer verbally on most tasks, although they sometimes responded non-verbally. However, there was one mute participant who answered all questions non-verbally. In addition, for the ToM and rule-use tasks, if participants failed to respond on a test question, the question was repeated in a forced-choice manner (e.g. 'Where is the kitten really, in the living room or in the kitchen?').
4. Some researchers opt to exclude data from participants who fail the reality question. However, although most participants fail ToM tasks by reporting current reality (i.e. a realist bias), studies that analyse data from both questions (e.g. Charman *et al.*, 1998; Frye *et al.*, 1995; Zelazo *et al.*, 1996) show that some children fail the tasks by making phenomenist errors on reality questions. Excluding children who err on the reality question leads to an overestimation of the prevalence of a reality bias. It also eliminates those children who by chance get the reality question correct and the ToM question wrong, which seems to assume that participants could only fail the ToM question because of poor reasoning about mental states (i.e. that the ToM question is a pure measure of mental state understanding).
5. The correlation between ToM and rule-use composites remains high if we assign scores of zero to participants who failed the reality question, and give participants who passed the reality question one point for each mental state question passed ($r = 0.78$, $p < 0.01$). Moreover, the correlation does not change if we include only those participants with a clear diagnosis (i.e. using data from 9 of the 10 mildly impaired participants; $r = 0.80$, $p < 0.01$).
6. Each of the rule-use measures was significantly correlated with the ToM composite for the mildly impaired group ($r = 0.80$, $p < 0.01$, for the DCCS; $r = 0.68$, $p < 0.05$, for the ramp task). Likewise, performance on both the unexpected-contents ToM tasks and the explicit false-belief task was significantly correlated with performance on the rule-use composite ($r = 0.79$, $p < 0.01$, for the unexpected-contents ToM tasks; $r = 0.62$, $p < 0.06$, for the explicit false-belief task).
7. As was the case for the mildly impaired group, the correlation between rule-use and ToM performance does not change when we assign scores of zero to participants who failed the reality question, and give participants who passed the reality question one point for each mental state question passed, $r = -0.40$, $p > 0.10$, or when only those participants with clear diagnoses are included ($n = 10$), $r = -0.33$, $p > 0.10$.

REFERENCES

American Psychiatric Association. 1987. *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., revised). Author: Washington, DC.

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders* (4th edn, text revision). Author: Washington, DC.

Anderson SW, Damasio AR, Damasio H. 1990. Troubled letters but not numbers: Domain specific cognitive impairments following focal damage in frontal cortex. *Brain* **113** (Pt 3): 749–766.

Atran S. 1998. Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences* **21**: 547–609.

Barkow JH, Cosmides L, Tooby J (eds). 1992. *The Adapted Mind*. Oxford University Press: Oxford.

Baron-Cohen S. 1994. How to build a baby that can read minds: Cognitive mechanisms in mind reading. *Current Psychology of Cognition/Cahiers de Psychologie Cognitive* **13**: 513–552.

Baron-Cohen S. 1995. *Mindblindness*. MIT Press: Cambridge, MA.

Baron-Cohen S, Leslie AM, Frith U. 1985. Does the autistic child have a 'theory of mind'? *Cognition* **21**: 37–46.

Baron-Cohen S, Leslie AM, Frith U. 1986. Mechanical, behavioural, and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology* **4**: 113–125.

Blair, RJR. 1995. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition* **57**: 1–29.

Burack JA. 1997. The study of atypical and typical populations in developmental psychopathology: The quest for a common science. In *Developmental Psychopathology: Perspectives on Adjustment, Risk, and Disorder*, Luthar SS, Burack JA, Cicchetti D, Weisz JR (eds), Cambridge University Press: New York; 139–165.

Burack JA, Zigler E. 1990. Intentional and incidental memory in organically mentally retarded, familial retarded, and nonretarded individuals. *American Journal on Mental Retardation* **94**: 532–540.

Caramazza A, Shelton JR. 1998. Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience*. **10**: 1–34.

Carlson SM, Moses L. 2001. Individual differences in inhibitory control and children's theory of mind. *Child Development* **72**: 1032–1053.

Charman T, Campbell A, Edwards LS. 1998. Theory of mind performance in children, adolescents, and adults with mental handicap. *Cognitive Development* **13**: 307–322.

Cicchetti D. 1984. The emergence of developmental psychopathology. *Child Development* **55**: 1–7.

Damasio A. 1990. Category–related recognition defects as a clue to the neural substrates of knowledge. *Trends in Neuroscience* **13**: 95–98.

Das Gupta P, Bryant PE. 1989. Young children's causal inferences. *Child Development* **60**: 1138–1146.

Dunn LM, Dunn LM. 1981. PPVT: *Peabody picture vocabulary test–revised: Manual for forms L and M*. American Guidance Services: Circle Pines, MN.

Dunn J, Kirsner K. 1988. Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review* **95**: 91–101.

Elman JL, Bates E, Johnson MH, Karmiloff-Smith A, Parisi D, Plunkett K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press: Cambridge, MA.

Flavell JH, Flavell ER, Green FL. 1983. Development of the appearance-reality distinction. *Cognitive Psychology* **15**: 95–120.

Fodor JA. 1983. *Modularity of Mind*. MIT Press: Cambridge, MA.

Frith U. 1989. *Autism: Explaining the Enigma*. Basil Blackwell: Oxford.

Frith U, Happé F. 1994. Autism: Beyond ''theory of mind.'' *Cognition* **50**: 115–132.

Frye D, Zelazo PD (in press). The development of young children's action control and awareness. In *Agency and Self-Awareness: Issues in Philosophy and Psychology*, Roessler J, Eilan N (eds). Oxford University Press: Oxford.

Frye D, Zelazo PD, Brooks PJ, Samuels MC. 1996. Inference and action in early causal reasoning. *Developmental Psychology* **32**: 120–131.

Frye D, Zelazo PD, Burack JA. 1998. I. Cognitive complexity and control: Implications for theory of mind in typical and atypical development. *Current Directions in Psychological Science* **7**: 116–121.

Frye D, Zelazo PD, Palfai T. 1995. Theory of mind and rule-based reasoning. *Cognitive Development* **10**: 483–527.

Gopnik A, Astington JW. 1988. Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development* **59**: 26–37.

Gopnik A, Wellman HM. 1994. The theory theory. In *Mapping the Mind: Domain Specificity in Cognition and Culture*, Hirschfeld LA, Gelman SA (eds). Cambridge University Press: New York; 257–293.

Gordon A, Olson D. 1998. The relation between acquisition of a theory of mind and information processing capacity. *Journal of Experimental Child Psychology* **68**: 70–83.

Griffith EM, Pennington BF, Wehner EA, Rogers SJ. 1999. Executive functions in young children with autism. *Child Development* **70**: 817–832.

Hagen J. 1967. The effect of distraction on selective attention. *Child Development* **38**: 685–694.

Halford G, Wilson WH, Phillips S. 1998. Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences* **21**: 803–864.

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

Happé FGE, Frith U. 1995. Theory of mind and autism. In *Learning and Cognition in Autism, Current Issues in Autism,* Schopler E, Mesibov IGB (eds). Plenum Press: New York; 177–197.

Harris PL. 1993. Pretending and planning. In *Understanding Other Minds: Perspectives From Autism*, Baron-Cohen S, Tager-Flusberg H, Cohen DJ (eds), Oxford University Press: Oxford; 228–246.

Hermelin B, O'Connor N. 1970. *Psychological Experiments with Autistic Children*. Pergamon Press: Oxford.

Hodapp RM, Burack JA. 1990. What mental retardation teaches us about typical development: The examples of sequences, rates, and cross-domain relations. *Development and Psychopathology* **2**: 213–225.

Hughes C. 1998. Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology* **16**: 233–253.

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

Hughes C, Russell J, Robbins TW. 1994. Evidence for executive dysfunction in autism. *Neuropsychologia* **32**: 477–492.

Jacques S, Zelazo PD. 1994. Several strictures on specificity. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition* **13**: 623–629.

Karmiloff-Smith A. 1992. *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press: Cambridge, MA.

Leekam SR, Perner J. 1991. Does the autistic child have a metarepresentational deficit? *Cognition* **40**: 203–218.

Leevers HJ, Harris PL. 2000. Counterfactual syllogistic reasoning in normal 4-year-olds, children with learning disabilities, and children with autism. *Journal of Experimental Child Psychology* **76**: 64–87.

Leslie AM. 1994. ToMM, ToBy, and Agency: Core architecture and domain specificity. In *Mapping the Mind: Domain Specificity in Cognition and Culture*, Hirschfeld LA, Gelman SA (eds). Cambridge University Press: New York; 119–148.

Leslie AM, Thaiss L. 1992. Domain Specificity in Conceptual Development: Neuropsychological Evidence from Autism. *Cognition* **43**: 225–251.

Luthar SS, Burack JA, Cicchetti D, Weisz JR. (eds). 1997. *Developmental Psychopathology: Perspectives on Adjustment, Risk, and Disorder*. Cambridge University Press: New York.

Marr D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman: San Francisco.

McEvoy RE, Rogers SJ, Pennington BF. 1993. Executive function and social communication deficits in young autistic children. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **34**: 563–578.

Oswald DP, Ollendick TH. 1989. Role taking and social competence in autism and mental retardation. *Journal of Autism and Developmental Disorders* **19**: 119–138.

Ozonoff S. 1995. Executive functions in autism. In Schopler E, Mesibov GB (eds), *Learning and Cognition in Autism. Current Issues in Autism* Plenum Press: New York; 199–219.

Ozonoff S, Pennington BF, Rogers SJ. 1991. Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **32**: 1081–1105.

Perner J. 2000. About+belief+counterfactual. In *Children's reasoning and the mind*, Mitchell P, Riggs K. (eds). Psychology Press: Hove, UK; 367–401.
Perner J, Lang B. 1999. Development of theory of mind and executive control. *Trends in Cognitive Sciences* **3**: 337–344.
Perner J, Leekam SR, Wimmer H. 1987. Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology* **5**: 125–137.
Perner J, Stummer S, Lang B. 1999. Executive functions and theory of mind: Cognitive complexity or functional dependence? In *Developing Theories of Intention: Social Understanding and Self-Control*, Zelazo PD, Astington JW, Olson DR (eds). Lawrence Erlbaum: Mahwah, NJ; 133–152.
Peterson DM, Bowler DM. 2000. Counterfactual reasoning and false belief understanding in children with autism, children with severe learning difficulties, and children with typical development. *Autism* **4**: 391–405.
Piaget J. 1954. *The Construction of Reality in the Child*. Basic Books: New York.
Prior MR. 1979. Cognitive abilities and disabilities in infantile autism: A review. *Journal of Abnormal Child Psychology* **7**: 357–380.
Prior MR, Dahlstrom B, Squires T.-L. 1990. Autistic children's knowledge of thinking and feeling states in other people. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **31**: 587–602.
Prior MR, Hoffman W. 1990. Brief report: Neuropsychological testing of autistic children through an exploration with frontal lobe tests. *Journal of Autism and Developmental Disorders* **20**: 581–590.
Quartz S, Sejnowski TJ. 1994. Beyond modularity: Neural evidence for constructivist principles in development. *Behavioral and Brain Sciences* **17**: 725–726.
Quartz S, Sejnowski TJ. 1997. The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences* **20**: 537–596.
Riggs KJ, Peterson DM. 2000. Counterfactual thinking in pre-school children: Mental state and causal inferences. In *Children's reasoning and the mind*, Mitchell P, Riggs K (eds). Psychology Press: Hove, UK; 87–99.
Roth D, Leslie AM. 1998. Solving belief problems: Toward a task analysis. *Cognition* **66**: 1–31.
Russell J, Mauthner N, Sharpe S, Tidswell T. 1991. The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology* **9**: 331–349.
Schank RC, Ableson RP. 1978. *Scripts, Goals, Plans and Understanding*. Lawrence Erlbaum: Hillsdale, NJ.
Scott FJ, Baron-Cohen S. 1996. Logical, analogical, and psychological reasoning in autism: A test of Cosmides' theory. *Development and Psychopathology* **8**: 235–245.
Scott FJ, Baron-Cohen S, Leslie A. 1999. 'If pigs could fly': A test of counterfactual reasoning and pretence in children with autism. *British Journal of Developmental Psychology* **17**: 349–362.
Slaughter V. 1998. Children's understanding of pictorial and mental representations. *Child Development* **69**: 321–332.
Sperber D. 1994. The modularity of thought and the epidemiology of representations. In *Mapping the mind: Domain specificity in cognition and culture*, Hirschfeld L, Gelman SA (eds). Cambridge University Press: New York; 39–67.
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.
Treisman AM, Kanwisher NG. 1998. Perceiving visually presented objects: Recognition, awareness, and modularity. *Current Opinion in Neurobiology* **8**: 218–226.
Wellman H, Bartsch K. 1988. Young children's reasoning about beliefs. *Cognition* **30**: 239–277.
Wimmer H, Perner J. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13**: 103–128.
Yirmiya N, Erel O, Shaked M, Solomonica-Levi D. 1998. Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin* **124**: 283–307.
Zelazo PD, Burack J, Benedetto E, Frye D. 1996a. Theory of mind and rule use in individuals with Down Syndrome: A test of the uniqueness and specificity claims. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **37**: 479–484.

Zelazo PD, Burack J, Boseovski J, Jacques S, Frye D. 2001. A cognitive complexity and control framework for the study of autism. In *The development of Autism: Perspectives from Theory and Research*, Burack JA, Charman T, Yirmiya N, Zelazo PR (eds). Lawrence Erlbaum Associates: Mahwah, NJ; 195–217.
Zelazo PD, Carter A, Reznick JS, Frye D. 1997. Early development of executive function: A problem-solving framework. *Review of General Psychology* **1**: 198–226.
Zelazo PD, Frye D. 1997. Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In *Language Structure, Discourse, and the Access to Consciousness*, Stamenov M. (ed). John Benjamins: Amsterdam; 113–153.
Zelazo PD, Frye D. 1998. II. Cognitive complexity and control: the development of executive function. *Current Directions in Psychological Science* **7**: 121–126.
Zelazo PD, Frye D, Rapus T. 1996b. An age-related dissociation between knowing rules and using them. *Cognitive Development* **11**: 37–63.
Zelazo PD, Helwig CC, Lau A. 1996c. Intention, act and outcome in behavioral prediction and moral judgement. *Child Development* **67**: 2478–2492.
Zelazo PD, Jacques S. 1996. Children's rule use: Representation, reflection and cognitive control. In *Annals of Child Development*, vol. 12, Vasta R (ed). Jessica Kingsley: London; 119–176.
Zelazo PD, Mueller U. in press. Executive function in typical and atypical development. In *Handbook of Childhood Cognitive Development*, Goswami U. (ed). Blackwell: Oxford.
Zelazo PD, Reznick JS, Spinazzola J. 1998. Representational flexibility and response control in a multistep multilocation search task. *Developmental Psychology* **34**: 203–214.
Zigler E, Hodapp RM. 1986. *Understanding Mental Retardation*. Cambridge University Press: New York.

